2008

Solution Conformations and Structural Thermodynamics of Type I DNA Polymerases from Escherichia coli and Thermus aquaticus

Allison Joubert Richard
Louisiana State University and Agricultural and Mechanical College

SOLUTION CONFORMATIONS AND STRUCTURAL THERMODYNAMICS OF TYPE I DNA POLYMERASES FROM *ESCHERICHIA COLI* AND *THERMUS AQUATICUS*

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Biological Sciences

by
Allison Joubert Richard
B.S., Louisiana State University, 2001
May 2008

DEDICATION

This dissertation is partially dedicated to my parents, Michael and Sandy Joubert. They have always taught me that with God anything is possible, and many times when my research and writing seemed quite impossible, I was reminded of their lesson and motivated to push through the difficult times.

Above all, I dedicate this dissertation to my beloved husband, Briar Richard. He is my support system, my rock, my partner, and my greatest friend. His dedication to his own intellectual pursuits has always motivated me to succeed in this challenging endeavor. I am sincerely greatful for his endless love and support.

ACKNOWLEDGEMENTS

It is with my deepest and sincere gratitude that I thank my advisor, Dr. Vince LiCata. I would not have completed this dissertation if it were not for his extreme patience, editing, and encouragement. His advice as both a mentor and a friend has been invaluable. Thank you, Vince, for believing in my abilities as a scientist (even when I sometimes do not). Also, his passion for integrating science and art has always been an inspiration to me.

This work would not have been possible without the invaluable assistance of many wonderful people. Thanks to Iris Torriani, Tomas Plivelic, Cristiano de Oliveira, and John Pople who first taught me the technique of small angle X-ray scattering at the LNLS and SSRL synchrotrons. I gratefully acknowledge Marc Niebuhr, Thomas Weiss, and Hiro Tsuruta, beamline scientists at SSRL, for their support during our data collection runs and for all of their hard work in transforming the biological SAXS beamline into a much more user-friendly and automated machine. Thanks to Dr. Dmitri Svergun and all the members of his research group for helpful discussions regarding the analysis of my X-ray scattering data. The X-ray scattering-based shape modeling presented in this dissertation was performed using the High Performance Computing (HPC) resources at Louisiana State University. I am extremely thankful to Shangli Ou from the Center for Computation and Technology (CCT) and Dr. Randall Hall from the LSU Chemistry Department for their instruction on using the HPC resources. Dr. Alexandra Klinger hosted me as an intern at Johnson and Johnson PRD, where I was able to perform the Klenow stability study using the ThermoFluor technology. I am grateful for this experience.

I would like to acknowledge all of the members of the LiCata Lab, past and present, who have provided a wonderful and friendly work environment. I am especially indebted to all of those, including my advisor, who have helped me with protein preparations and to collect X-ray

scattering data into the wee hours of the morning: Carmen Ruiz, Andy Wowor, Chin-Chi (Kitty) Liu, Greg Thompson, Allyn Schoeffler, Thomas Cleveland, and Kausiki Datta. Thanks also to Angela Byrd who taught me analytical ultracentrifugation (AU) and who helped to collect much of the hydrodynamic data. I am especially grateful to Daniel Deredge and Carmen Ruiz who are always willing to lend an ear whether it be for scientific or personal discussion; your friendship means a lot to me and has definitely helped to keep me sane.

I extend my gratitude to all those serving on my dissertation committee: Drs. Jackie Stephens, Marcia Newcomer, Anne Grove, and Thomas Gillis.

My dissertation research using small angle X-ray scattering certainly would not have been possible without the financial support of the NSF and the NSF-IGERT training grant fellowship #998703. Dr. Paul Russo has tirelessly administered this grant, and I am much indebted to his efforts.

Last but certainly not least, I would like to thank my family and friends for their love, support, and encouragement; especially, my mother-in-law, Mary Richard, whose love and prayers have gotten me through some of my most difficult moments in the completion of this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The global conformations of DNA polymerase I from *Thermus aquaticus* (Taq/Klentaq) and *E. coli* (Pol I/Klenow) both in isolation and in complex with DNA have been examined in solution using hydrodynamic (analytical ultracentrifugation) and small/wide angle X-ray scattering techniques and then compared to their known crystal structures to assess the similarities and differences in the overall structure of these enzymes and enzyme complexes within a solution environment. These studies address the orientation of the 5' nuclease domain with respect the polymerase domain (elongated or compact) for the full-length polymerases, and the binding mode ("polymerization" versus "editing") of the DNA substrate when bound to Klentaq and Klenow. Comparisons of experimental and structure-based data indicate that full-length Pol I and Taq in solution adopt a conformation where the 5' nuclease domain is extended away from the polymerase domain, similar to the elongated crystal structure. *Ab initio* shape models of the apo polymerases generated from the scattering data demonstrate remarkable likeness to their corresponding crystal structures and also reveal regions of flexibility. For DNA bound Klenow and Klentaq, comparative analyses indicate that 1) the global conformations of the complexes are not dependent upon the structure of the DNA substrate (primed-template versus blunt-ended) but are polymerase specific, 2) DNA binds to Klenow in the editing mode and to Klentaq in the polymerization mode, and 3) the solution structures deviate somewhat from the crystal structure-based models.

Additionally, the stability landscape of Klenow, as monitored by high-throughput thermal denaturation in a variety of solution conditions, demonstrates that Klenow's melting temperature (Tm) increases with increasing salt concentration and decreasing pH; Klenow's Tm spans from 40 to 62^{o}C. Both cation and anion specific stabilization is observed. The cationic stabilization

of Klenow can be well explained by a model postulating dampening of repulsion within surface anionic patches on the protein.

Both the global conformation and the stability studies demonstrate the importance of the solution environment in the comprehensive characterization of an enzyme's structure and function. The ability to visualize these polymerases and polymerase complexes in solution promises to open new avenues of understanding of these important enzymes.

CHAPTER 1

GENERAL INTRODUCTION

1.1 *E. coli* Pol I

One essential attribute of all living organisms is the ability to reproduce and transfer genetic information from parent to offspring. The enzymes primarily responsible for faithfully conserving the flow of genetic information throughout future generations are the DNA polymerases. Beginning in the mid-1950's and over the course of the following 15 years, Arthur Kornberg and colleagues discovered, isolated, and characterized the first DNA polymerase, the type I DNA polymerase from *Escherichia coli* (Pol I) [1-6]. The major structural and functional attributes first characterized with Pol I, and its ability to synthesize DNA *in vitro*, remain the conceptual framework for almost all current understanding of nucleic acid biosynthesis [4-6]. The experiments of Kornberg and associates on *E. coli* DNA Pol I revealed the following properties that are now known to be shared by all DNA polymerases: 1) catalysis of nucleotide incorporation is template-directed and guided by geometric interactions of Watson-Crick base pairs within the polymerase active site; 2) elongation of the nascent DNA strand proceeds in the 5' → 3' direction, anti-parallel to the template chain; and 3) phosphodiester bond formation by DNA polymerases requires an RNA or DNA oligonucleotide primer with a free terminal 3' hydroxyl (OH), (i.e. DNA polymerases are not capable of initiating DNA synthesis *de novo)* [6, 7].

Since the discovery of *E. coli* Pol I, many other DNA polymerases from all three domains of life (eubacteria, archaea, and eukarya) and various bacteriophages and viruses have been identified and characterized [8]. However, Pol I persists as a model enzyme for the study of polymerase function during DNA replication.

1.2 Other Polymerase Families

DNA polymerases are collectively characterized into a superfamily based on their functional ability to catalyze the formation of a phosphodiester bond between the terminal nucleotide of a growing DNA chain and an incoming deoxynucleotide triphosphate (dNTP) in a template dependent fashion [6]. Amino acid sequence homology [9-11] and crystal structure comparisons [12-14] reveal that the DNA polymerase superfamily can be broken down into seven distinct families. In addition to the polymerase core, DNA polymerases typically possess additional family specific enzymatic activities or specialized protein/nucleic acid interaction surfaces that reside within distinct accessory domains [8].

Most organisms have evolved multiple, specialized DNA polymerases to efficiently manage genome replication and a variety of repair processes. For many years Pol I was thought to be *E. coli*'s sole polymerase. In addition to Pol I, *E. coli* have at least four other distinct DNA polymerases, designated as types II, III, IV, and V (numbered consecutively in order of discovery) [7]. Initially the activity of Pol I, which is typically the most abundant *E. coli* polymerase, masked the presence of the other DNA polymerases [15]. An *E. coli* cell contains ~400 molecules of Pol I and only 10 – 40 type III polymerase particles [7, 16]. However, the discovery of viable *E. coli* expressing Pol I mutants, deficient in polymerase activity [15], suggested the existence of additional *E. coli* DNA polymerases, and eventually led to the identification of Pol III as the primary replicative DNA synthesizing enzyme of *E. coli.* Polymerases II, IV, and V are important in various repair processes [7]. *E. coli* Pol I is essential for both the replacement of Okazaki fragment primers following lagging strand synthesis, and repair of damaged DNA [6, 15].

While the enzymatic activity of polymerases developed presumably during the most primitive evolutionary stages of life, currently seven distinct DNA polymerase families are recognized. Based on amino acid sequence similarity [9-11] and structural homology [12], the known DNA polymerases from all three domains of life (eubacteria, archaea, and eukarya), as well as viral and bacteriophage DNA polymerases are grouped into families A, B, C, D, X, Y, and RT [8-12]. The innately complex processes of DNA metabolism (a collective term for DNA replication, repair, and recombination) are even more elaborate in archaea and eukaryotes. However, all DNA polymerases seem to share a fundamental "two-metal-ion" molecular mechanism for catalyzing nucleotide incorporation [17]. The family association of each DNA polymerase is not restricted by its domain-of-life affiliation. For example, 1) the five eubacterial polymerases (designated by the roman numerals I, II, III, IV, and V) are the representative members of families A (Pol I), B (Pol II), C (Pol III), and Y (Pols IV and V), 2) families B and Y include some DNA polymerases from all three domains of life, and 3) bacterial Pol III and certain archaeal polymerases from the Euryarchaeota subdomain (family D), which share no sequence similarity with other DNA polymerases, each constitute a distinct family [8]. Two family A DNA polymerases, *E. coli* Pol I and *Thermus aquaticus* Taq, are the focus of this dissertation; therefore, discussion of other DNA polymerases is limited here to a general overview of each family and edifying comparative analyses.

Some DNA polymerases are specialized for genome replication, whereas others are dedicated to DNA repair processes. Thus, for many of these enzymes in addition to the 5' → 3' polymerase active site, other active sites with supplemental catalytic activities are present within distinct, separable domains of the same polypeptide. These auxiliary enzymatic activities include 5' → 3' nuclease activity (family A), 3' → 5' exonuclease activity (families A, B, and

D), lyase activity (X family), and RNaseH activity (RT family) [8]. While some polymerases, such as the type I eubacterial DNA polymerases, function alone as a single polypeptide chain, many others, especially the primary replicative DNA synthesizing enzymes, function *in vivo* as holoenzymes composed of many accessory subunits, which endow the holoenzyme with additional structural framework, extra protein or DNA interaction surfaces, and a variety of other enzymatic activities. For example, processivity factors are common, important accessory subunits. These factors promote continuous nucleotide incorporation by decreasing the ability of the polymerase to dissociate from the DNA substrate [7].

Based on homology with *E. coli* DNA Pol I, all eubacterial type I DNA polymerases belong to family A. Other members of this family include eukaryotic mitochondrial DNA polymerase γ and the replicative polymerases of bacteriophages T3, T5, and T7 [8]. The eubacterial type I DNA polymerases are not the primary replicative polymerases in bacteria. They mainly function in repair and in processing of Okazaki fragments following lagging strand synthesis during replication [6]. While *E. coli* Pol I is the best characterized member of this family, Taq DNA polymerase, from *Thermus aquaticus*, is the second most well studied eubacterial family A polymerase. The properties of Pol I and Taq are often extrapolated between species to highlight polymerase features that are assumed to be common among all family A DNA polymerases. These comparisons often extend to members of other DNA polymerase families to identify general and specific polymerase characteristics.

Family B includes the eukaryotic DNA bypass polymerase zeta (ζ) and the major replicative eukaryotic DNA polymerases α, δ, and ε, as well as several DNA synthesizing enzymes belonging to archaea, viruses, bacteria (Pol II), and bacteriophages. While the major mitochondrial DNA polymerase of eukaryotes (pol γ) is a member of family A, other plasmid

derived mitochondrial polymerases found in plants and fungi belong to family B. Some of the most well studied members of this family are the bacteriophage DNA polymerases RB69, T4, and Phi29 [8, 18]. For the most part, family B enzymes are the primary replicative polymerases in their cell types (except for within the eubacteria), thus members of this family are typically processive and thus can incorporate 10-20 thousands of nucleotides per binding event [7, 8]. The majority of the known B family polymerases exhibit superior 3' → 5' exonuclease activity, which helps to ensure faithful DNA replication [8].

The principal replicative bacterial DNA synthesizing enzymes do not share significant sequence homology with any other polymerase and are thus classified alone in family C. These enzymes associate with many other proteins and function as a holoenzyme. The subunit containing the DNA polymerase activity is, by convention, denoted the α subunit. *E. coli* Pol III is the best characterized member of this family. The Pol III holoenzyme is composed of 10 different subunits. The majority of these subunits are processivity factors [7, 8]. Along with α, the ε and θ subunits make up the catalytic core of Pol III. The ε subunit has 3' → 5' exonuclease that is stimulated by the θ subunit [19]. The first crystal structures of the Pol III α subunit from *E. coli* [20] and *T. aquaticus* [21] were only recently solved in 2006. Both of these structures demonstrate that while the catalytic core of the bacterial replicative polymerase shares a hand-like architecture that is common to all known polymerases (the details of this structural scaffold are discussed further below), the structural topology of the catalytic center (palm subdomain) more resembles that of the X family polymerase β rather than the eukaryotic and archaeal replicative polymerases [20, 21].

X and Y family polymerases are typically involved in various repair processes. Eukaryotic polymerases β, σ, μ, and λ, yeast polymerase IV, and the African swine fever virus X

polymerase are classified as X family members. Pol β plays a role in base excision repair (BER) and is the most well studied member of this family [8].

1.3 Structure and Function of Pol I and Taq

Taq DNA polymerase is the type I DNA polymerase from the thermophilic eubacterium, *Thermus aquaticus*. This enzyme is structurally and functionally homologous to *E. coli* Pol I but has evolved to function in an environment that is approximately 40 – 60°C hotter than the optimal physiological growth temperature of *E. coli* (37°C) [16, 22-24]. Taq and Pol I share about 38% sequence identity [25]. While the polymerase activity of Taq has a temperature optimum of 75°C, partial activity and structural stability are retained up to temperatures between 95 and 100°C [23, 26]. The extreme thermostability of Taq makes it a particularly useful biotechnological reagent in the polymerase chain reaction (PCR), which has become a commercially valuable technique because of its ability to greatly amplify small quantities of DNA [23, 27].

Full-length *E. coli* Pol I is a single polypeptide chain with three distinct structure-function domains: a polymerization domain, a 3' exonuclease (proofreading) domain, and a 5' nuclease domain. Removal of the 5' nuclease domain yields the Klenow "large fragment", which is a fully functional DNA polymerase [28, 29]. Taq shares an almost identical architecture [23]. The functional activities are organized into three structural domains, demonstrated by the full-length Taq DNA polymerase crystal structure [30] shown in Figure 1.1.

Template-directed DNA synthesis is catalyzed within the polymerase domain. The 3' exonuclease domain, which is inactive in Taq, is the "proofreading" domain. It scans newly replicated DNA for base pairing errors and excises misincorporated nucleotides from the 3' end of the nascent DNA strand during replication. The presence of an active 3' exonuclease domain

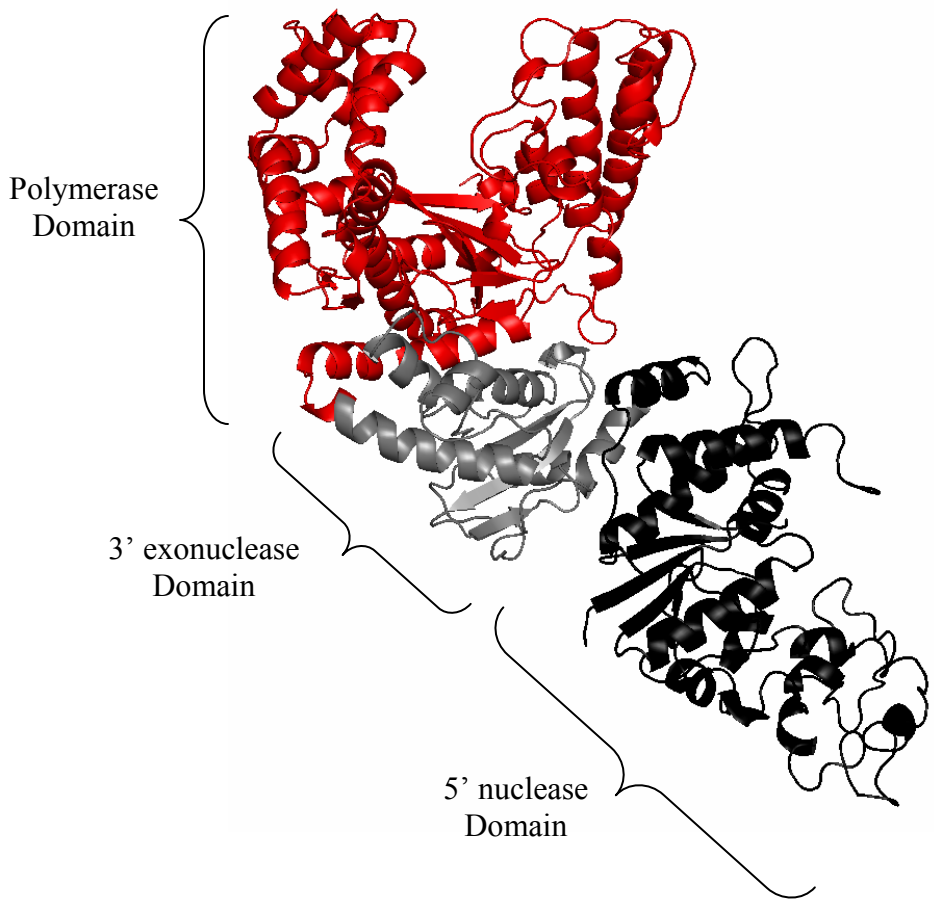


Figure 1.1. Structure-function domains of Taq DNA polymerase (PDB ID: 1TAQ). The 3’ exonuclease activity is non-functional in Taq polymerase.

significantly increases replication fidelity [6]. Taq lacks 3’ exonuclease activity because its respective domain is missing the essential carboxylate-functionalized residues required for catalytic activity [23]. The 5’ nuclease domain is primarily responsible for the removal of RNA primers, which are laid down at the beginning of each Okazaki fragment during lagging strand synthesis. Damaged DNA nucleotides are also removed by the action of the 5’ nuclease domain during nucleotide excision repair (NER) [6, 15, 31, 32]. The complete repair of damaged DNA and the Okazaki fragments requires the coordinated activity of the polymerase and 5’ nuclease domains (a topic which is specifically addressed in Chapter 4). Since the polymerase initiates

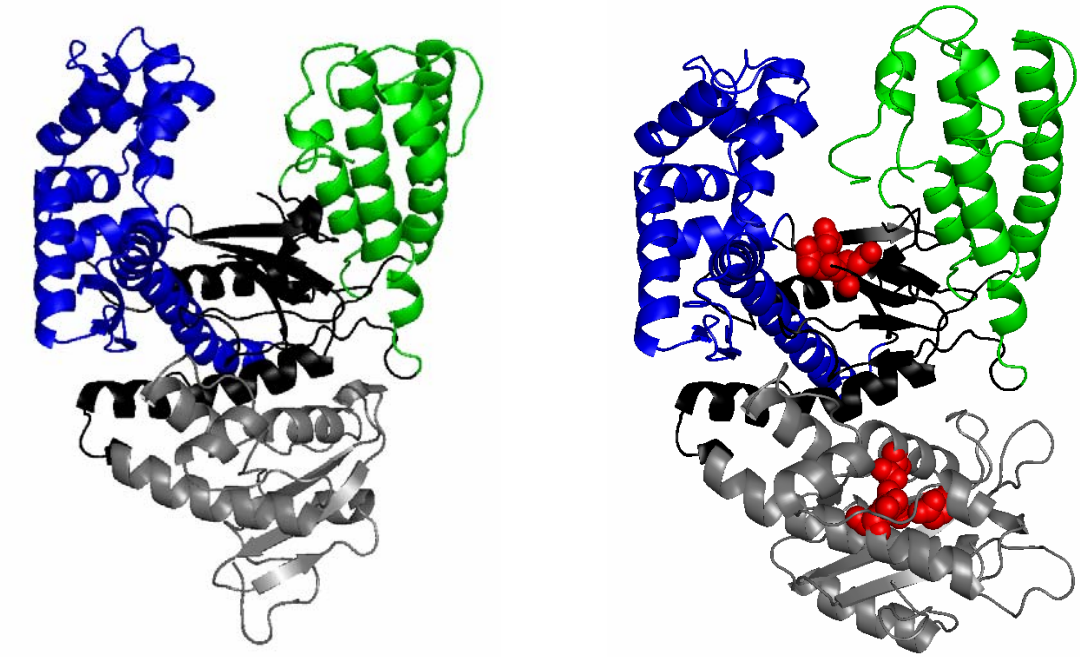
repair at a nick or single-stranded gap in the DNA duplex during such repair, and then terminates with a new nick somewhere downstream, this process in known as ”nick translation”. According to several studies of a variety of *E. coli* Pol I mutants, some of which were temperature-sensitive or conditionally lethal, it was determined that only the 5’ nuclease activity of Pol I is essential for *E. coli* viability, but that both the polymerase and 5’ nuclease activities are required for efficient nick translation (reviewed in [15]).

Of the known type I DNA polymerases, the only full-length structural data comes from crystal structures of Taq. As mentioned, in nick-translation the 5’ nuclease and polymerase domains must work in coordination to yield a final single-stranded nick that is ligatable by DNA ligase. Crystal structures of full-length Taq demonstrate both an elongated (Figure 1.1) and a compact configuration of this polymerase’s functional domains that are dependent upon the orientation of the 5’ nuclease domain with respect to the Klentaq fragment. These two distinct conformations (which are discussed in more detail in Chapter 4 of this dissertation) have diverse functional implications. The X-ray scattering data in Chapters 4 and 5 reveal that in solution apo full-length Taq adopts an elongated global conformation that is extremely similar to the conformation shown in Figure 1.1.

Removal of the 5’ nuclease domain from Taq polymerase leaves the Klenow fragment equivalent, which is denoted “Klentaq”. Klenow and Klentaq share nearly 50% amino acid sequence identity between their polymerase domains [23, 33], and their respective crystal structures [33-35] show that these two proteins are highly structurally homologous (see Figure 1.2). The polymerase domain adopts a shape that resembles a half-opened right hand with “fingers,” “thumb,” and “palm” subdomains. The palm subdomain contains the catalytic residues. All polymerases seem to share the hand-like architectural topology of the polymerase

domain. The structure of the palm subdomain is relatively conserved within the DNA polymerase superfamily, while the detailed structural features of the fingers and thumb subdomains differ between families. Despite a lack of significant primary sequence homology between DNA polymerases from different families, the elucidation of three-dimensional structures of polymerases from various families has demonstrated a strong conservation of this topological structure [12, 14].

Crystal structures of Klenow, Klentaq, and full-length Taq DNA polymerases have been determined both in the absence and presence of substrate. The majority of the studies in this dissertation focus on comparisons of these crystal structures with solution conformations. As



Klentaq (1KTQ) **Klenow (1KRP)**

Figure 1.2. Structural similarity between Klentaq and Klenow DNA polymerases. In both structures, the polymerase domain is color-coded to emphasize its hand-like architecture: fingers (blue), palm (black), and thumb (green). The 3' proofreading domains (inactive in Klentaq) are shown in gray. For the Klenow crystal structure, the polymerase and 3' exonuclease catalytic residues are colored red and shown in a space-fill representation to highlight the distance between these active sites.

noted above, the polymerase domain of both proteins has been likened to a half-opened right hand with a fingers, thumb, and palm subdomain. The fingers subdomain is responsible for binding dNTP and delivering this substrate to the active site. The thumb subdomain makes direct interactions with the DNA duplex that are likely important in processivity [36]. While the fingers and thumb are mostly α-helical, the palm subdomain is comprised of an antiparallel β-sheet that contains the active site residues [35] (see Figure 1.2).

Comparative analyses between the Klenow [35, 37, 38] and Klentaq [33] crystal structures were integral in understanding the molecular-basis for Klentaq's proofreading inactivity. These two proteins differ mainly in their 3' exonuclease domains, which despite a lack of sequence homology [23] appear to share a significant level of similarity in the topology of their secondary structural features, albeit several of the helicies and loops in Klenow are altered, shortened, or completely missing in Klentaq. The overall size of Klentaq's vestigial proofreading domain is smaller than that of Klenow [33] (see Figure 1.2). Most importantly, in Klentaq all of the 3' exonuclease active site carboxylate containing residues are missing or mutated, along with many of the residues important in binding of the single-stranded DNA substrate. The binding site for the primer terminus is filled in by hydrophobic side chains, and thus it is unlikely that Klentaq's 3' exonuclease domain retains the ability to bind single-stranded DNA [33].

1.4 Recent Studies of the DNA Binding Topologies by Klenow and Klentaq

A number of questions about the structural details of the interactions of Klentaq and Klenow with their DNA substrates are the focus of the bulk of the studies of this dissertation, as well as the subject of a number of recent studies in other laboratories. These include: 1) In what binding mode does the DNA initially bind to each polymerase? 2) How are the details of DNA

binding topology different or similar for the two polymerases? 3) How does the DNA shuttle between the polymerase and proofreading active sites in Klenow? 4) Does the initial binding mode depend on the structure of the DNA substrate? 5) Do the polymerases always bind as monomers at the ends of DNA fragments? and 6) How do solution structures of the polymerases and polymerase-DNA complexes compare to the known crystal structures of these enzymes? For each of these questions, the existing data from different techniques and different laboratories conflict somewhat. These issues are discussed briefly in the remaining subsection of this Introduction.

Two DNA binding modes, a "polymerization" mode and an "editing" mode, have been described for the type I DNA polymerases based on the current collection of family A DNA polymerase crystal structures, which include DNA bound co-crystal structures of Klentaq [39, 40], Klenow [37], full-length Taq [41], *Bacillus stearothermophilus (Bst)* Pol I [42], and the T7 replicative DNA polymerase [43]. In these structures, Klentaq, Taq, *Bst* Pol I, and T7 DNA polymerase all bind DNA in the polymerization mode, while only Klenow demonstrates editing mode binding; DNA binding in both modes has not been structurally observed for any of these enzymes. Like Klentaq/Taq, *Bst* Pol I is thermophilic and lacks proofreading activity [44]. Conversely, T7 DNA polymerase is a replicative polymerase and actually exhibits 3' exonuclease activity that is greater than 100-fold higher than that of Klenow [45]. If presence of a functional editing site correlates with editing-mode initial binding, it is obvious to ask why T7 DNA polymerase was not crystallized with its DNA bound in the proofreading domain. However, the single-stranded binding cleft observed in Klenow is blocked in T7, and Doublie et al. suggest that this obstruction may result from structural rearrangements due to a six-residue deletion engineered to deactivate the enzyme's 3' exonucleolytic activity [43]. Since all of the

polymerization mode complexes exhibit similar features, specific details of the polymerization and editing DNA binding modes will only be further discussed in the context of Klentaq and Klenow.

Originally, based on the first Klenow crystal structure [35], which did not contain DNA, the double-stranded portion of the DNA substrate had been modeled as binding within the polymerase cleft between the fingers and thumb subdomains. This seemed to be the obvious binding site for the DNA duplex seeing as its width and depth were both slightly larger than 20 Å, which is the diameter of B-form DNA [35]. According to this model and consistent with protection results from footprinting studies [46] and primer photoaffinity crosslinking analyses [47], the polymerase interacts with approximately 8 base pairs (bps) of duplex DNA when the 3' primer terminus is positioned near the proposed catalytic active site. However, the model resulted in slight steric overlap between the protein and nucleic acid at several sites within narrower regions of the cleft [35]. Subsequent observations suggesting that the catalytic active site was deeper into the cleft than originally proposed induced further steric conflicts and could not be reconciled with the independently determined binding site footprint [37, 48]. The editing mode Klenow-DNA duplex co-crystal structure [37] revealed for the first time the now commonly accepted binding site for the duplex product, which lies in the cleft between the thumb and 3' exonuclease domain.

Subsequently, the polymerase cleft was proposed to bind the 5' single-stranded portion of template [37]; however, none of the crystal structures directly support this hypothesis. Both crystallographic [39, 42, 43] and crosslinking [49] data suggest that the single-stranded template strand binds across the fingers subdomain. Although the cleft surface residues are highly conserved between the family A DNA polymerases [44, 50], and they show a positive

electrostatic potential (likely to be important for binding negatively charged DNA) in Klenow, to date no functional significance has been directly attributed to this distinctive polymerase cleft. Perhaps, the cleft results purely as a consequence of the need for the functional mobility of the fingers and thumb subdomains.

In both the polymerization and editing DNA binding modes (see Figure 1.3) the DNA duplex binds in the same region between the thumb and the proofreading domain. However, the angle at which the DNA exits the polymerase differs between the two binding modes, and the helical axis of the DNA is translocated by ~3-4 base pairs [41]. Polymerization mode binding positions the 3' terminus of the primer strand near the catalytic residues within the polymerase

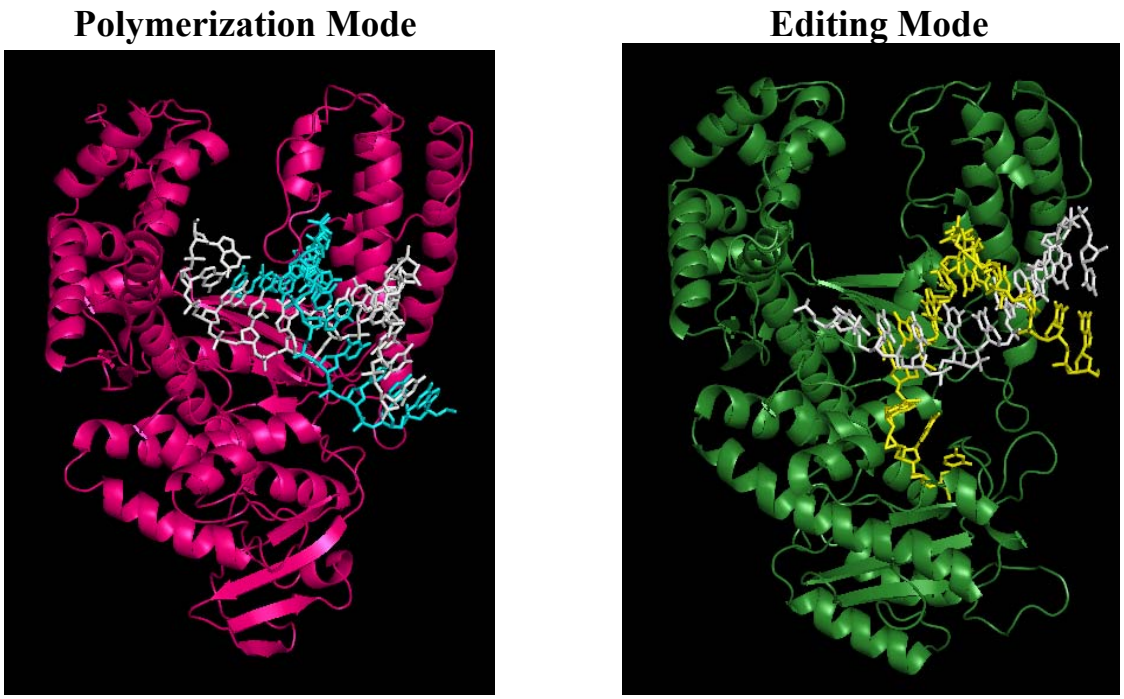


Figure 1.3. Polymerization and editing DNA binding modes for Klentaq and Klenow. *Left:* Klentaq bound to DNA in the polymerization mode. *Right:* Klenow complexed with DNA in the editing mode. The single-stranded 3' overhang interacts with the proofreading domain, such that the terminal nucleotide is positioned near the 3' exonuclease active site. The DNA template strand in both structures is colored white, while the DNA primer strands are colored cyan and yellow, respectively, in the 4KTQ and 1KLN structures.

domain. Conversely, in the Klenow editing complex, 3-4 nucleotides at the 3' end of the primer strand are melted away from the other strand of the DNA duplex and are bound within a narrow cleft within the proofreading domain [37]. Using crosslinked duplex DNA substrates, Cowart and associates confirm that at least 4 bps of duplex DNA are melted in editing mode binding [51]. As described in detail in Chapter 6 of this dissertation, X-ray scattering solution-based studies suggest that 4 bp is indeed a lower limit, and that 6 – 8 bp may be melted apart during initial binding in the editing mode to Klenow.

For Klenow DNA polymerase, the polymerase and proofreading active sites are separated by 30-35 Å (see Figure 1.2). A rapid intramolecular shuttling mechanism has been proposed to describe the transfer of the 3' primer terminus between the polymerase and editing active sites during processive nucleotide incorporation. The proofreading domain binds a single-stranded (ss) 3' terminus, while the polymerase binds a duplex primer terminus. According to the DNA bound Taq/KTQ and KLN crystal structures, the duplex portion of the DNAs bind in a similar position between the thumb and 3' exonuclease domains, but their helical axes are translated relative to one another [41]. Thus, it is possible for the primer terminus to dissociate from one active site, the polymerase to slide along the DNA without dissociating, and the terminus to rebind at the alternative active site. Experimentally, Joyce has shown, using DNA trapping experiments to monitor a single processive binding event, that transfer of the 3' primer terminus for Klenow can proceed via both intramolecular and intermolecular processes, and that a mismatched substrate favors intermolecular transfer [52].

Millar and associates have suggested that even for correctly base-paired nucleotides, the 3' exonuclease domain binds the ss primer terminus ~14% of the time [53], indicating an equilibrium between the two active sites, in which the rate of transfer is faster than either

catalysis by the polymerase or 3' exonuclease. Mismatches increase melting of the DNA duplex and stalling of the polymerase activity, which consequently both shift the equilibrium to the proofreading site [53]. Interestingly, Millar and associates have also recently proposed that Klenow can bind to DNA as a dimer [54]. The X-ray scattering data in this dissertation strongly refute this hypothesis, and no other data on Klenow since the 1950's has suggested that this may be the case.

Traditional methods of determining DNA binding thermodynamics are steady-state fluorescence, fluorescence anisotropy, and electrophoretic mobility shift assays. However, none of these techniques can dissect the thermodynamic contributions of polymerization and editing mode binding on the total binding affinity. The major direct insights into partitioning of initial binding between the two sites have thus come from: 1) a time-resolved anisotropic method developed by Millar and associates [53, 55, 56], 2) a conformationally sensitive CD-spectrophotometric assay developed by von Hippel and associates that can detect whether DNA is in single-stranded or double-stranded form while bound to a protein [57, 58], and the X-ray scattering studies of this dissertation. Experiments by Millar and associates estimate that ~14% of the 3' ends partition into the editing domain for perfectly matched DNA. The fraction of 3' ends binding in the editing domain steadily increases as the number of mismatches increases up to 4 mismatches, at which point nearly 100% of the ends are bound in the editing mode. When the total equilibrium constant for DNA binding to Klenow is dissected into its polymerization and editing mode contributions, it is evident that the mismatches decrease the binding affinity to the polymerase domain, but do not alter the affinity of the 3' exonuclease for its single-stranded substrate [53]. In contrast to this scenario, data from von Hippel and associates, using the 2-aminopurine dimer spectrophotometric assay indicate that perfectly matched primed-template

(pt) DNA binds to Klenow in editing mode essentially 100% of the time (K. Datta and P.H. von Hippel, personal communication). Likewise, the data of this dissertation indicate that for Klenow, the initial binding of matched ptDNA is primarily in the editing mode.

According to studies from both Joyce [52] and Millar and associates [53, 55], the structure of the DNA substrate (matched versus mismatched) significantly affects the partitioning of the primer terminus between the polymerase and exonuclease binding sites of Klenow. In our laboratory, using traditional fluorescence anisotropy measurements, Klenow exhibits differences in binding affinity (and other thermodynamic binding parameters) for matched primed-template versus blunt-ended DNA, while Klentaq demonstrates identical binding affinity for both DNA structures (Wowor and LiCata, unpublished). These results may reflect an altered partitioning equilibrium between Klenow's polymerase and exonuclease sites for binding of these two DNA structures, which would not be expected for Klentaq (since its ss binding cleft within the exonuclease domain is shown to be blocked [33]). According to the X-ray scattering results in Chapter 6 of this dissertation, Klenow binds both DNA structures primarily in the editing mode, while Klentaq binds both structures in the polymerization mode. Thus, if binding site partitioning in Klenow is differentially affected by the alternate DNA structures, then it is not significant enough switch the predominant binding mode from editing to polymerization.

1.5 Solution-based Global Conformations and Structure-based Thermodynamic Stability

Understanding the conformation and structure-based thermodynamic stability (structural thermodynamics) of an enzyme or enzyme-substrate complex in solution is an essential aspect of the comprehensive characterization of the overall structure and function of that enzyme. The studies in this dissertation address the 1) stability landscape of Klenow DNA polymerase in

solution, 2) the spatial relationship between the 5' nuclease domain and Klentaq fragment of full-length Taq DNA polymerase, 3) solution conformations of apo Taq, Klentaq, and Klenow, including regions of flexibility that can be observed by X-ray scattering, and 4) the global solution conformations of Klentaq and Klenow DNA polymerases in complex with DNA. In all of these studies, pertinent questions of biological mechanism, as summarized at the beginning of subsection 1.4 and in 1.3, are addressed. The final question discussed at the beginning of the previous subsection: "How do solution structures of the polymerases and polymerase-DNA complexes compare to the known crystal structures of these enzymes?" is addressed extensively in Chapters 5 and 6 of this dissertation. The ability to visualize these polymerases and polymerase complexes in solution, which is the main accomplishment of this dissertation, promises to open new avenues of understanding of these important enzymes. Before proceeding to the hydrodynamic and X-ray scattering studies of the type I DNA polymerases from *T. aquaticus* and *E. coli*, in Chapter 2 the thermal stability landscape of Klenow is characterized as a function of pH, salt concentration, and salt type as well as a variety of other solution conditions, and the observed cationic stabilization of Klenow is considered in the context of the electrostatic potential topology of Klenow's surface.

CHAPTER 2

THERMAL STABILITY LANDSCAPE FOR KLENOW DNA POLYMERASE AS A FUNCTION OF pH AND SALT CONCENTRATION[A]

2.1 Introduction

Solution conditions have long been known to influence protein stability. There are a small number of broad spectrum salt plus pH surveys of the thermal stability landscapes for different proteins in the scientific literature, and they demonstrate that the effects of salt and pH on protein stability are highly protein specific [59-64]. In other words, pH changes and salt additions can either increase or decrease stability, depending on the protein, and their effects are not necessarily parallel. Numerous studies of the effects of salt on protein stability exist, and for the majority of proteins examined to date, added salt stabilizes the protein [60-72]. This effect is common enough that deviations from it are notable for their rarity (e.g. [59]). Elucidation of the underlying molecular mechanisms for salt-induced stabilization of proteins, however, is still an active area of research. The responses of a specific protein to salt or pH changes are strong reflections of the nature of the protein itself, and such characterizations can reveal what types of non-covalent interactions and surface properties are linked to that protein's stability.

We have previously characterized the thermal and chemical denaturation of Klenow polymerase in directly comparative studies with the thermophilic Pol I Type DNA polymerase from *Thermus aquaticus* [26, 73]. Those studies revealed extremely large differences in Tm

---

[A] Reprinted from Biochimica et Biophysica Acta, 1764, A.J. Richard, C.C. Liu, A.L. Klinger, M.J. Todd, T.M. Mezzasalma, and V.J. LiCata, Thermal stability landscape for Klenow DNA polymerise as a function of pH and salt concentration, 1546-1552, Copyright 2006, with permission from Elsevier.

and stabilizing free energies (ΔG) between the partners in this mesophilic-thermophilic protein pair, and characterized some of the differences in the unfolding processes for the two polymerase species. In order to facilitate direct comparison, the characterizations in these previous studies were performed under identical conditions for both polymerases.

In this study we have expanded our understanding of the stability of Klenow DNA polymerase by surveying the effects of salt concentration, specific ions, and pH. Joyce and associates also previously measured Tm values for Klenow and several site directed mutants using CD spectroscopy in a solution of 20 mM potassium phosphate, 15 mM NaCl, and 15% glycerol at pH 7.0 [74]. The Tm values obtained in the studies of Joyce and associates were significantly higher (by 8 - 15^{o}C) than those obtained in our previous studies. The results of the present study clearly show that the differences between Joyce's Tm measurements and our own Tm measurements are explained by the pH-salt stability landscape.

Klenow denatures irreversibly, thus precluding thermodynamic analysis of its denaturation. Most proteins denature irreversibly [75]. Despite this, characterization of the thermal and chemical denaturation of irreversibly denaturing proteins can lead to a variety of new information at the molecular level, and frequently can add to understanding the biology and physiology of the protein.

This study utilizes a miniaturized fluorescence assay that measures Tm based on an increase in the binding of 1,8-anilinonaphthalene sulfonate (ANS) to the denatured state of the protein [76-78]. We find that the Tm for Klenow in this assay ranges from 40 to 62^{o}C, depending on the solvent conditions, with both increases in salt and decreases in pH leading to greater stabilization. Our data also indicate that there are strong ion-specific cation effects on stability, overlaid onto a more typical Hofmeister anion type dependence. Examination of the

surface electrostatic topology of Klenow has shown that the majority of the surface of the protein has a net negative charge [79]. Attenuation of surface anionic repulsion thus appears to be a reasonable explanation for the cationic stabilization of Klenow.

2.2 Materials and Methods

2.2.1 Materials

Klenow DNA polymerase was overexpressed and purified according to previously described procedures [80, 81]. Following purification, the protein was stored at -20°C in 50 mM Tris, 0.5 mM DTT, and 50% glycerol, pH 7.5 until use.

2.2.2 ThermoFluor Measurements

The ThermoFluor®[B] thermal shift assay system at Johnson & Johnson Pharmaceutical Research & Development was used to measure the thermal stability of Klenow DNA polymerase under a wide variety of buffer, pH, and salt conditions. This assay has been described in detail previously [76, 77]. For reversibly unfolding proteins, this assay has been shown to be as thermodynamically reliable as conventional methods [76]. For irreversibly unfolding proteins, such as Klenow, only measured Tm values are used to deduce information about the protein. For the present study, 2 μl of each experimental buffer solution, with all components at two times the desired assay concentration, were dispensed into black 384-well polypropylene PCR microplates (Abgene), along with 2 μl of the 2X protein solution, which included the ANS fluorescent probe. Each assay well contained 0.075 mg/ml Klenow in a final volume of 4 μl. Different assay wells

---

[B] The ThermoFluor assay was developed by 3-Dimensional Pharmaceuticals, Inc., which has been merged into Johnson & Johnson Pharmaceutical Research & Development, L.L.C. "ThermoFluor" is a trademark registered in the United States and certain other countries.

contained different buffers, with different salts, at different pH values, as described in the appropriate figure legends.

Buffers used included 25 mM sodium acetate, MES, sodium phosphate, PIPES, HEPES, MOPS, and sodium borate (used to generate pH's from 4 – 9). Salts included 5 – 505 mM NaCl, 0 and 5 mM $MgCl_2$, 100 and 300 mM NH_4Cl, LiCl, and KCl, 10 and 100 mM $(NH_4)_2SO_4$, (NH_4)-PO_4, $MgSO_4$, K-PO_4, $LiSO_4$, and $NiSO_4$, and 5 mM $MnCl_2$. Each assay well also contained 0.1 mM ANS, 1 mM Tris, 2.5 mM Pipes, 0.01 mM DTT, 1% glycerol, and 0.5% DMSO. To prevent evaporation during thermal denaturation, 1 μl of silicone oil (Fluka) was dispensed over the assay solution in each well. Assay plates were then loaded onto a heating block within the ThermoFluor machine. Resident data acquisition programs were used to control the run and collection parameters. The following parameters were used for all experiments: 1) samples were heated from 25 to 85°C, 2) samples were held for 15 seconds at each 1°C increment, and imaged for 10 seconds during the hold time, 3) the gain was set to 2, 4) the imaging time was 10 seconds, 5) filtered UV light at 380 – 400 nm was used to excite the fluorescent probe, and 6) the fluorescence emission intensity was measured at 500 ± 25 nm using a CCD camera to simultaneously collect an intensity measurement for each of the 384 wells. For each well, the intensity was the sum of the intensity detected by all pixels within the area of a box that was 8 pixels by 8 pixels.

The Tm, or midpoint of the thermal transition, can be determined either by using the maximal value of the derivative of the transition region of the melting curve, or by fitting to extended form of the van't Hoff relationship, as described previously [76]. In the case of reversibly denaturing proteins, true thermodynamic parameters, such as the ΔH of unfolding, may be obtained from the data [76]. In irreversible systems, such as Klenow, only the Tm values

are determined. Requisite controls, such as establishing that the ANS probe does not interact with the native state of the protein, have also been described previously [76, 77].

2.2.3 Circular Dichroism (CD) Measurements

For CD measurements at different pH values, spectra from 218-230 nm were recorded using an AVIV model 202 circular dichroism spectrophotometer. The protein concentration was 0.1 mg/ml, and measurements were made in 10 mM phosphate buffer with 2 mM β-mercaptoethanol. pH was adjusted by mixing monobasic, dibasic, and tribasic phosphate buffers. For CD monitored thermal denaturation in the presence and absence of $NiSO_4$, protein concentration was 0.05 mg/ml in 10 mM Tris, pH 7.9. $NiSO_4$ was 10 mM. Data were collected from 25 to 60°C in a rectangular cuvette (1 cm path length) with a screw top seal and constant stirring. The CD signal at 222 nm was fitted to a modified form of the van't Hoff equation, which simultaneously fits the native and denatured baselines and the transition region to obtain the Tm for denaturation, as described previously [81].

2.2.4 Electrostatic Potential Contour

GRASP was used to calculate the electrostatic potential surface map for Klenow DNA polymerase [82]. Coordinates for Klenow (1KFD) were obtained from the Protein Data Bank. Full charges were assigned. Adjustable parameters used were default values, except that ionic strength was set at 50 mM.

2.3 Results

Figure 2.1 shows typical denaturation curves obtained for Klenow polymerase in this study. In this figure, denaturation curves as a function of increasing NaCl concentration are shown. Data are obtained with a miniaturized fluorescence-based assay performed using the

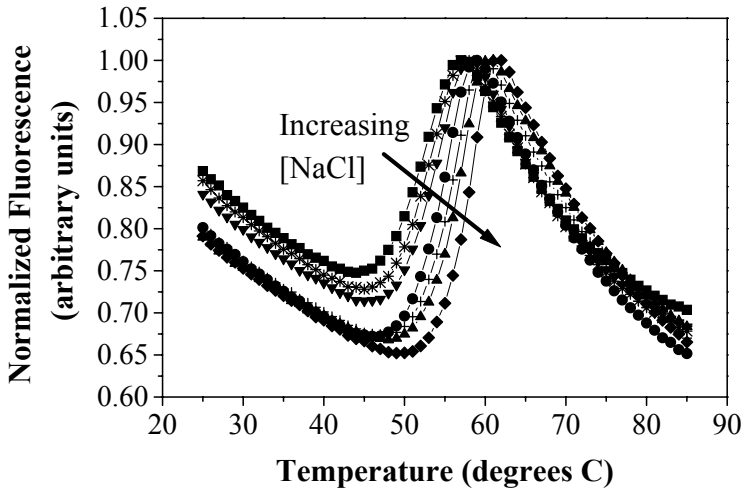


Figure 2.1. Thermal melting curves for Klenow polymerase. Each individual curve, normalized for maximum fluorescence intensity, shows the thermal denaturation of Klenow at a different NaCl concentration in 25 mM phosphate buffer, pH 7.0. The NaCl concentrations are 0 mM (squares), 50 mM (stars), 100 mM (triangles), 200 mM (circles), 300 mM (crosses), 400 mM (inverted triangles), and 500 mM (diamonds). Tm is determined, as usual, from the midpoint of the unfolding transition.

ThermoFluor® assay system at Johnson & Johnson Pharmaceutical Research and Development [76, 77]. In this assay, protein unfolding is monitored by following the increase in fluorescence due to the enhanced binding of 1, 8-anilinonaphthalene sulfonate (ANS) to the denatured state of the protein. The denaturation curves consist of the same three segments as more traditional denaturation curves: an initial native baseline, a transition region where the protein unfolds, and a subsequent denatured state baseline. The native and denatured state baselines for a "ThermoFluor" denaturation have more significant slopes than with many other denaturation methods due to the temperature dependence of ANS fluorescence. Controls are performed to insure that ANS does not interact with the native state of the protein (data not shown). As with any thermal denaturation curve, the midpoint of the transition region provides the Tm value. With reversibly denatured proteins, the curves can also be analyzed for denaturation enthalpy and other thermodynamic parameters [76]. Klenow polymerase does not reversibly denature,

however, so our analyses here are restricted solely to the Tm. It should be noted that where it has been possible to directly compare Tm values, Thermofluor values average about 2°C higher than CD values. This is similar to typical variations seen between CD and calorimetric Tm values.

Simultaneous variation of both salt concentration and pH provides a stability surface or landscape for the protein, such as that shown in Figure 2.2 for Klenow. The pH by [NaCl] landscape in Figure 2.2 is, of course, only one of many such stability landscapes that one might examine for the protein. The use of different salts, or the inclusion or absence of "typical" protein chemistry additives (e.g. DTT, BME, EDTA, glycerol, surfactants, etc.) will potentially reveal different landscapes, or perhaps more correctly: a different transect through the global, multi-dimensional stability landscape for the protein. Furthermore, while we have focused only on Tm shifts, due to the irreversible denaturation of Klenow, stability landscapes for ΔG, ΔH, and ΔS versus solution conditions also exist and will be experimentally accessible for some proteins.

Figure 2.2 shows that increasing the concentration of NaCl at any pH will stabilize the protein. The magnitude of the stabilization changes at different pH's, giving rise to the topology of the stability surface. Likewise, decreasing pH at any salt concentration stabilizes the protein, but unlike the case for salt, there is a clear but subtle maximum in the pH stabilization, which gives the landscape its slightly peaked or hill-like appearance. Depending on the salt concentration, the pH of maximal stability hovers around pH 6.5-7.5. The decrease in stability in the pH 7 to 5 range is slight compared to the dramatic increases in stability in the pH 9 to 7 range, and so much of our discussion focuses on the more dramatic proton stabilization effect,

although it seems that an acid denaturation effect is likely starting to overlap at lower pH values. This onset of acid denaturation near pH 5.0 is also visible in Figure 2.3.

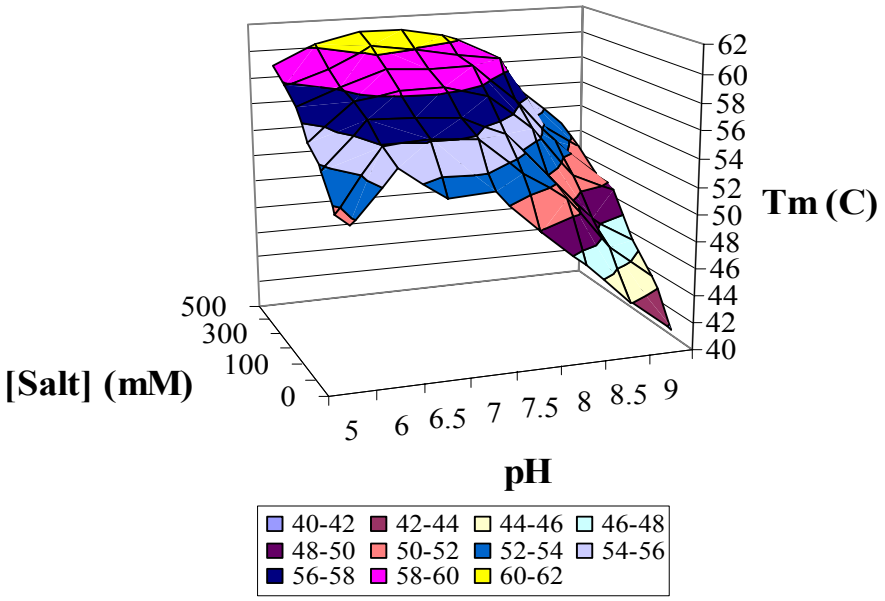


Figure 2.2. Thermal stability landscape for Klenow DNA polymerase as a function of pH and salt concentration. The different colors in the graph represent different Tm values in 3°C intervals from 40-62°C. Klenow stability increases with increasing salt concentration and for the most part with decreasing pH. Tm's at pH's 5-6.5 and 7.5-9 are averages of duplicates in the same buffer. Each pH 7 Tm is an average of the Tm for Klenow at the specified salt concentration in four different buffers (sodium phosphate, PIPES, HEPES, and MOPS).

Figure 2.3 is a plot of the circular dichroism of Klenow at 218-221 nm as a function of pH. In its native state, Klenow has a secondary structure trough in its CD spectrum in the 218-221 nm region, and this minimum is used to follow the thermal and/or chemical denaturation of the protein [26, 73]. This figure illustrates the transition zones for acid and alkali denaturation of the protein, which have midpoints around pH 4.5 and 10.4, respectively. These data show that with the possible exception of experiments near pH 5, that the secondary structure of Klenow is in the native state throughout the pH range examined in Figure 2.2. It is generally assumed that the protein is fully native under such conditions.

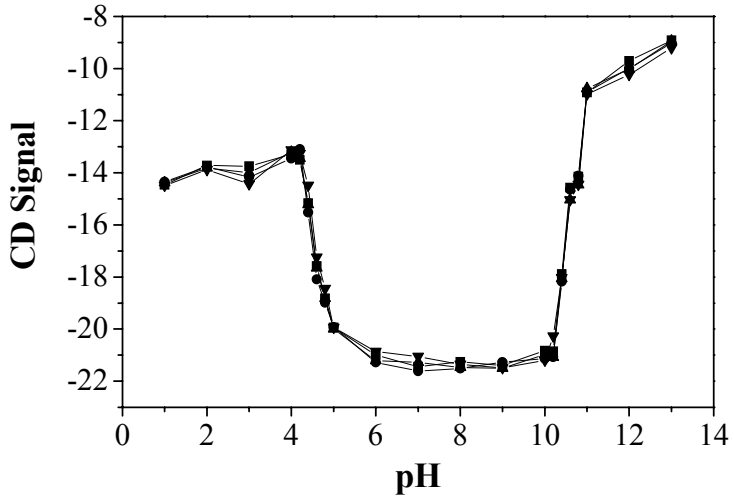


Figure 2.3. CD spectroscopy of Klenow DNA polymerase as a function of pH. This figure shows the CD signal (millidegrees) for Klenow monitored at 221 (squares), 220 (circles), 219 (triangles), and 218 nm (inverted triangles), measured as a function of pH.

Examining the stabilizing effects of salts other than NaCl reveals several interesting features about the stabilization of Klenow, and these are illustrated by the data in Figures 2.4-2.6. Figure 2.4 shows the ΔTm for additions of different concentrations of chloride salts (NaCl, KCl, NH_4Cl, $MgCl_2$, and $MnCl_2$), sulfate salts (Li_2SO_4, $(NH_4)_2SO_4$, $MgSO_4$, and $NiSO_4$), and phosphate salts (ammonium and potassium phosphate), as well as two common divalent metal chelators EDTA and EGTA. The particular salt and concentration combinations used are those of a typical "salt and pH screen" currently conducted as part of the drug screening and discovery procedures at Johnson & Johnson Pharmaceutical Research and Development [76, 77]. One can clearly observe both anionic and cationic effects in this figure. For example, for the four chloride salts of NH_4^+, Na^+, Li^+, and K^+, there is a clear hierarchy of cationic stabilization: $NH_4^+ > Na^+ > Li^+ > K^+$. Similarly, a comparison of sulfate or phosphate salts of the same cation, clearly show a 1-2 degree increase in Tm relative to chloride salts. The only salt that destabilizes Klenow is $NiSO_4$. In order to insure that the $NiSO_4$ destabilization observed in the ThermoFluor

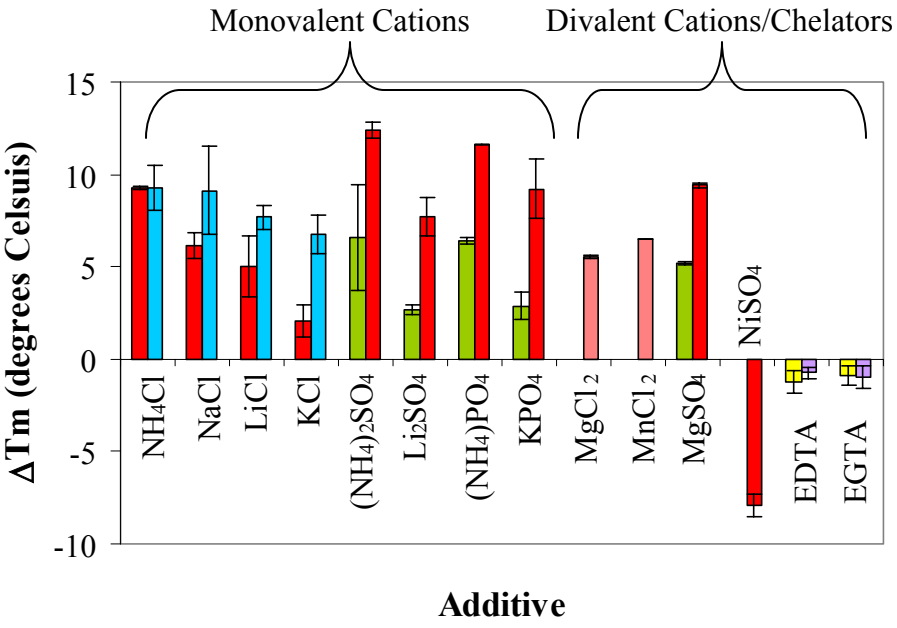


Figure 2.4. Thermal stabilization of Klenow is dependent upon both salt type and concentration. Each bar in this graph represents the change in Tm (ΔTm) for Klenow in the presence of the indicated additive with respect to a reference Tm (53.2 ± 0.5°C). The reference Tm is an average of 16 Tm's for Klenow diluted into water instead of additive. Error bars are the standard deviations of duplicate measurements. The bar shading is indicative of the additive concentration. Salt concentrations are: 5 mM (pink), 10 mM (green), 100 mM (red), and 300 mM (blue). Chelator concentrations are: 0.05 mM (yellow) and 1 mM (purple).

assay was not related to some peculiarity of the assay, we also thermally denatured Klenow in the presence of $NiSO_4$ in a circular dichroism monitored assay. Figure 2.5 thus confirms, using a more conventional assay, that the addition of $NiSO_4$ substantially lowers the Tm for Klenow.

Figure 2.6 shows another two-dimensional transect through the stability data for Klenow, the pH vs. Tm profile for Klenow under four specific combinations of NaCl and $MgCl_2$ (or their absence). Under all four conditions, increasing pH decreases the Tm, if all the data are linearly fit. The pH effect in the presence of 5mM $MgCl_2$, however, shows definite fine structure indicating there is a gentle (~1°C) maximum in this data. It is notable, however, that across most of the pH range adding 5 mM $MgCl_2$ has a significantly larger stabilizing effect than adding 100 mM NaCl. This again reinforces the presence of a significant cationic stabilization effect (since

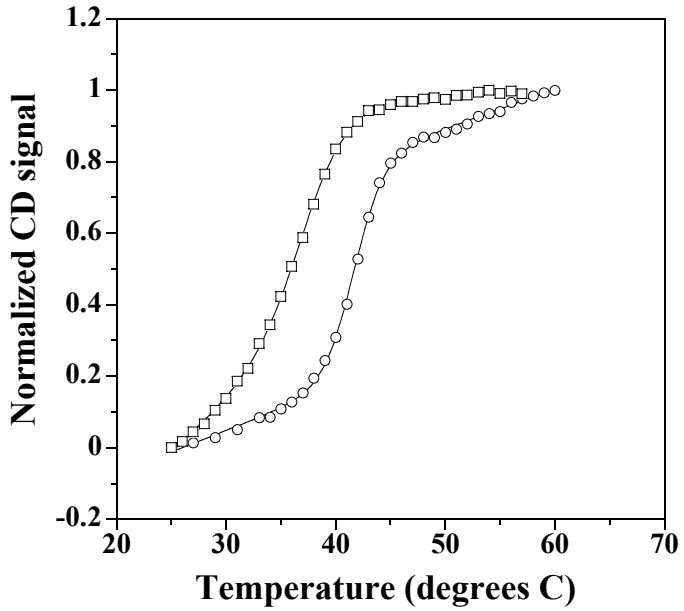


Figure 2.5. Circular dichroism monitored thermal denaturation of Klenow polymerase in the presence (squares) and absence (circles) of 10 mM $NiSO_4$. Protein concentration was 0.05 mg/ml in 10mM Tris, pH 7.9.

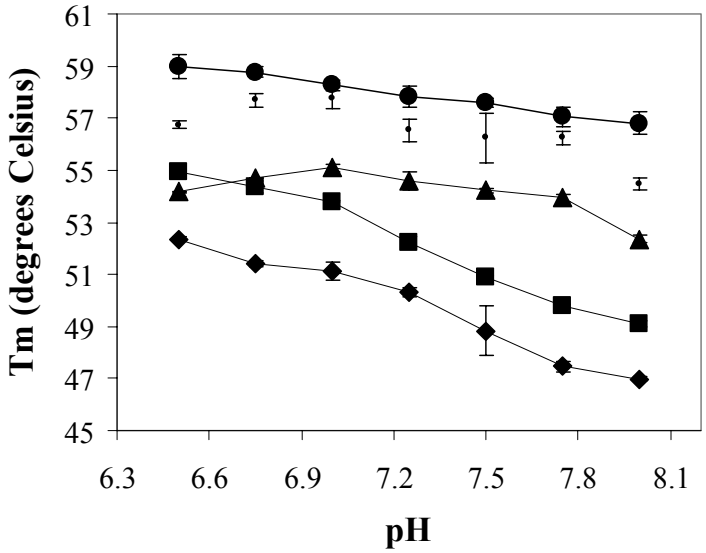


Figure 2.6. NaCl and $MgCl_2$ each individually increase the thermal stability of Klenow, and the combined stabilizing effect of these two salts is synergistic. The data sets show the dependence of Tm on pH for Klenow in 25 mM HEPES buffer with no added salt (filled diamonds), 100 mM NaCl (filled squares), 5 mM $MgCl_2$ (filled triangles), or 100 mM NaCl and 5 mM $MgCl_2$ (filled circles). Errors on all filled symbols are the standard deviations of two independent Tm measurements. The small dots represent the calculated additive stabilization of NaCl and $MgCl_2$, along with the associated propagated error.

there is only 10 mM Cl^- in 5 mM $MgCl_2$). It has long been known that Klenow binds Mg^{+2} (and Mn^{+2}) ions in its polymerization and exonuclease active sites [83, 84]. These Mg^{+2} ions are tightly bound, however, with Kd values in the micromolar concentration range [83, 84]. Even treatment of Klenow with 10 mM EDTA does not remove the Mg^{+2} ions necessary for DNA binding [81]. Nonetheless, the Mg^{+2} effect on Klenow stability may involve some overlap between specific active site bound Mg^{+2} uptake, and general surface Mg^{+2} association. Figure 2.6 also suggests that the effects of Mg^{+2} and Na^+ are synergistic. Across the entire pH range, but most significantly at high and low pH, the measured Tm in the presence of both salts is slightly higher than the sum of the measured ΔTm's for the two individual salts.

2.4 Discussion

In the context of a mutational characterization of the nucleotide binding site, Joyce and associates measured Tm values for Klenow and several site directed mutants of Klenow using CD spectroscopy in a solution of 20 mM potassium phosphate, 15 mM NaCl, and 15% glycerol at pH 7.0 [74]. Under these conditions, a Tm of 55°C was reported for wildtype Klenow [74]. These results were at odds with our own previous CD and calorimetric thermal denaturations of Klenow, which were conducted at high pH (9.5) and in the absence of salt in order to be directly comparable to parallel denaturations of Klentaq polymerase [26]. Under these no salt, high pH conditions we obtained a Tm of 37°C for Klenow, and demonstrated that decreasing the pH to 7.5 (still in the absence of salt) raised the Tm to 47°C [26]. The pH-salt stability landscape for Klenow determined in this study resolves the conflict between these previous measurements, and show that both Joyce's earlier Tm measurements and our own are easily explained by the pH-salt stability landscape for Klenow. Glycerol will also contribute somewhat to the previously observed differences in Tm for Klenow. We only examined Klenow in the presence of 1% and

5% glycerol (data not shown), and found that these additions caused 0.7 and 1.6°C increases in Tm, respectively.

2.4.1 Anionic and Cationic Stabilization

The typically stabilizing and occasionally destabilizing effects of salt on proteins has a long history of investigation [59-72]. Several molecular mechanisms have been proposed to explain the effects of salt on protein structure, including preferential hydration, specific ion binding, and electrostatic screening of repulsive surface charge interactions. Stabilizing and destabilizing effects of ions are most often attributed to the anion, and the Hofmeister series has long been known to be a general predictive/correlative guide for the stabilizing versus destabilizing effects of particular anions on protein structure, although a few exceptions exist. Chloride is consistently in the "middle" of the Hofmeister series, where the stabilizing and destabilizing effects are believed to be largely balanced, while sulfate and phosphate are consistently listed as two of the most stabilizing anions (although their exact order in the series sometimes differ from publication to publication). Figure 2.4 thus demonstrates clear Hofmeister effects on the stability of Klenow, since the sulfate and phosphate salts of any cation consistently show a 1-2°C stabilization relative to the chloride salt for that cation.

Anions are frequently found to be primarily responsible for the salt stabilization of proteins [62-64, 67]. Cation induced stabilization has also been observed, just less frequently [60, 68]. Figure 2.4 shows significant specific cation stabilization for Klenow polymerase. In chloride salts, the cationic stabilizing hierarchy is $NH_4^+ > Na^+ > Li^+ > K^+$. Fewer cation pairs were examined for phosphate and sulfate, but the stabilizing order to those that were examined is the same: $NH_4^+ > Li^+$ for sulfate and $NH_4^+ > K^+$ for phosphate.

2.4.2 Electrostatic Surface Topology

Previous electrostatic surface potential calculations performed for Klenow have shown that outside of the positively charged DNA binding crevice, the majority of the surface of the polymerase has a net negative charge [79]. The electrostatic surface of Klenow is shown in Figure 2.7. At physiological pH the protein will have a net negative charge of -7 to -8 (assuming each histidine residue contributes approximately +0.5 at neutral pH). Thus, the empirical stability behavior documented in this study can be well described by a model proposing that cations stabilize the protein by reducing the intrinsic repulsion within the predominantly anionic surface of the protein. Anionic dampening of surface cationic repulsion has been extensively studied by Fink and associates [63, 67, 71]. A similar mechanism for cationic stabilization, however, while previously hypothesized to be just as likely to exist [63], has proven somewhat empirically elusive [60]. Fink and associates have postulated that proteins that are stabilized by anionic attenuation of surface charge will tend to be proteins that are already marginally stable due to the surface cationic repulsion effects [63]. Klenow appears to follow this same model: we have previously shown that Klenow is very marginally stable for a protein of its size (68 kDa) [26, 73].

2.4.3 Destabilization by Nickel Sulfate

Figures 2.4 and 2.5 show an unusual, large destabilization of Klenow in the presence of $NiSO_4$. $NiSO_4$ destabilization of a number of proteins has been observed during the course of ThermoFluor drug screening at Johnson & Johnson Pharmaceutical Research and Development (data not shown). Since the vast majority of the proteins examined by ThermoFluor examined in this R&D program are his-tagged, it was generally assumed that the commonly observed $NiSO_4$ destabilization might originate from a preferential Ni^{+2}-histidine interaction in the denatured

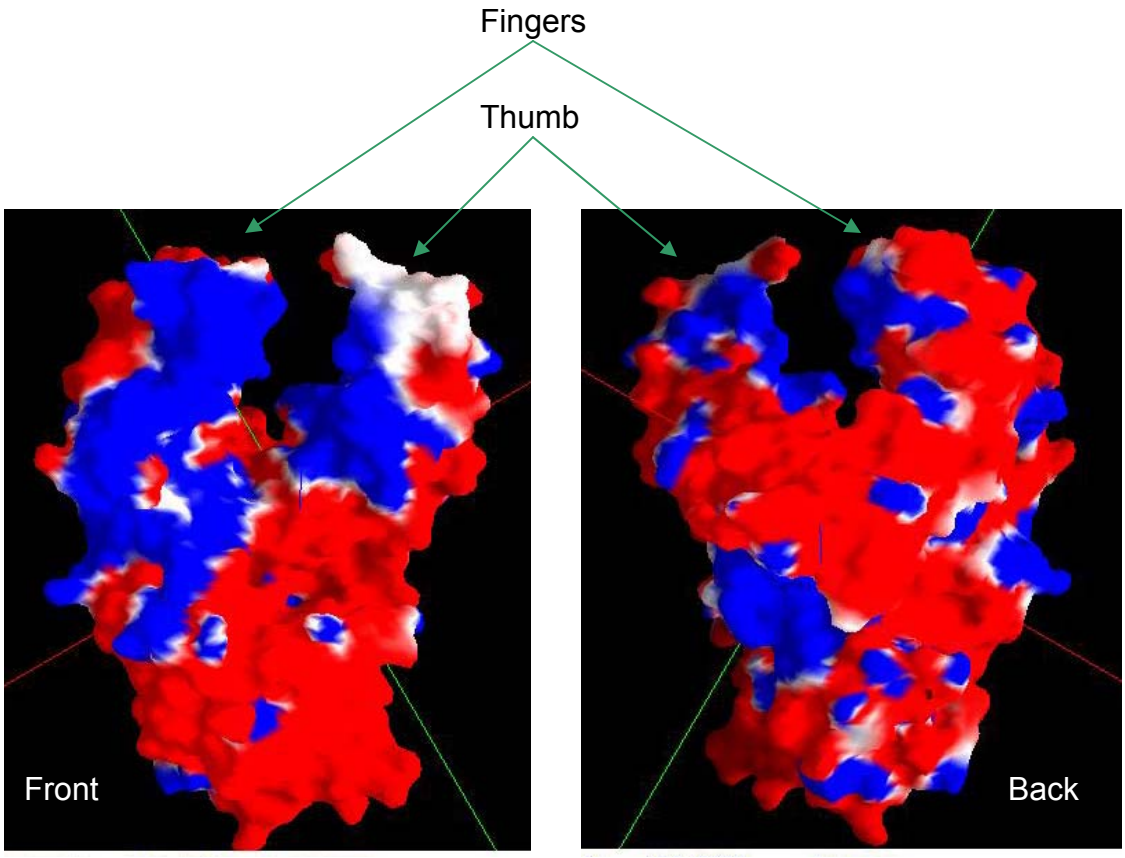


Figure 2.7. Topology of the electrostatic surface potential for Klenow DNA polymerase. Patches of positive (+) and negative (-) potential are shown, respectively, in blue and red. Neutral patches are white. The electrostatic topology of Klenow was originally calculated and published by Warwicker and associates [79]. The "fingers" and "thumb" domains of Klenow DNA polymerase are labeled in these diagrams to orient the viewer.

state, however the Klenow utilized in this study is not his-tagged. Ni^{+2} has been shown to generally stabilize human serum albumin and GroEL, and to specifically stabilize a variety of known Ni^{+2} binding proteins [85-88]. However, there is no evident prior documentation of a Ni^{+2} destabilizing effects on proteins, thus this large $NiSO_4$ destabilization effect may warrant further investigation. Our own preliminary examination of the effects of $NiSO_4$ on two other proteins found that this salt also destabilizes the adipocyte lipid binding protein, but stabilizes

lysozyme (data not shown). However, the extremely low solubility of nickel salts in many biological buffer systems seriously impedes further examination of these nickel effects.

2.4.4 Mg^{+2} versus Na^{+} Effects

Figure 2.6 shows the Tm values for Klenow 1) in the absence of added salt, 2) in the presence of 100 mM NaCl, 3) in the presence of 5 mM $MgCl_2$, and 4) in the presence of 100 mM NaCl + 5 mM $MgCl_2$. At all pH's except 6.5, 5 mM $MgCl_2$ has a larger stabilizing effect than 100 mM NaCl. If stabilization primarily occurs via ion interaction with the protein, then the affinity of Mg^{+2} for the protein is at least 10 fold higher than the affinity of Na^{+} for the protein. As discussed above (in Results) this may reflect some tight binding of Mg^{+2} to the active sites of the polymerase [83, 84]. The observation that combined Mg^{+2} and Na^{+} synergistically stabilize the protein to higher temperatures than the sum of their individual effects is also suggested by the data in Figure 2.6. Such synergy would require that Mg^{+2} and Na^{+} binding sites interact with each other, and act cooperatively to enhance the stability of the protein. For example, one might envision protein associated Mg^{+2} and Na^{+} ions being arranged relative to the protein surface in such a way that they might be bridged by Cl^{-} or water, leading to a synergistic stabilization. Such anion or water bridging could create a dynamic, semi-continuous, weakly bound ionic veil over the protein anionic patches.

There are several major caveats to the conceptual molecular models postulated in this Discussion. The first being that Tm need not necessarily be linearly additive like an enthalpy or free energy would be. Tm is a function of ΔG, ΔH, and ΔCp, but because the relationship is non-linear, perfectly linear combinations of ΔG's, ΔH's and ΔCp's could produce non-additive shifts in Tm values. Second, the irreversibility of unfolding means there will be some variability of Tm with rate of heating, but our previous studies have shown that this variability is minor

compared to the pH and salt induced changes characterized herein [26]. Further, the models described postulate that the cations interact directly with the protein, rather than exerting their effects indirectly, such as via preferential hydration effects [65, 70, 89]. Finally, we have assumed that the observed pH and salt effects are primarily a reflection of native state interactions, versus denatured state interactions. Quite simply, it is the striking correlation between anionic surface patching and the relatively uncommon cationic stabilization that makes cationic shielding of surface charge repulsion seem a particularly attractive explanatory model for the empirical pH-salt thermal stability landscape of Klenow.

CHAPTER 3

MATERIALS AND METHODS FOR CHARACTERIZING THE GLOBAL CONFORMATION OF MACROMOLECULES IN SOLUTION

3.1 Materials

3.1.1 Proteins

The proteins examined were the Klentaq fragment of Taq DNA polymerase [90, 91], the D424A "exo minus" mutant of Klenow [92], full-length Taq, full-length *E. coli* Pol I, His-tagged full-length Taq (His-Taq), and lysozyme from chicken egg white (Sigma-Aldrich, St. Louis, MO).

3.1.1.1 Polymerases without His-tag

Almost all experiments were performed with non-His-tagged proteins. Untagged Klenow, Klentaq, full-length Pol I, and full-length Taq were expressed and purified as described previously [81, 93]. No surfactants were used during purification, experiments, or storage of the polymerases. Taq and Klentaq polymerases were stored at 4^{o}C before use. Pol I and Klenow polymerases were stored frozen at -20^{o}C [80].

3.1.1.2 His-tagged Full-length Taq DNA Polymerase

3.1.1.2.1 Cloning

Datta and LiCata previously cloned the gene coding for full-length Taq DNA polymerase into a pTrc99A expression vector to create the pKDTaq2 plasmid [81], which unfortunately yielded only low levels of the Taq protein using the *E. coli* BL21(D3) expression system. Thus, the Taq gene was PCR-amplified from pKDTaq2, using aLIC4 and aLIC3 sense and anti-sense primers:

aLIC4 5'-GACGACGACAAGATGAATTCGGGGATGCTGCCCTC-3'

aLIC3 5'-GAGGAGAAGCCCGGTTCACTCCTTGGCGGAGAGCCAG-3',

and re-cloned into the pET-46 Ek/LIC *E. coli* T7 expression vector using an enterokinase (Ek)/ligation independent cloning (LIC) kit from Novagen (Madison, WI). As shown in Figure 3.1, the resulting plasmid, pAJR-Taq5, codes for Taq DNA polymerase tagged at the N-terminus with enterokinase cleavable hexahistidine (His-Taq). The calculated size of pAJR-Taq5, and the sequence and orientation of the Taq gene within the vector were confirmed by DNA sequencing and BamHI/HindIII restriction digestion.

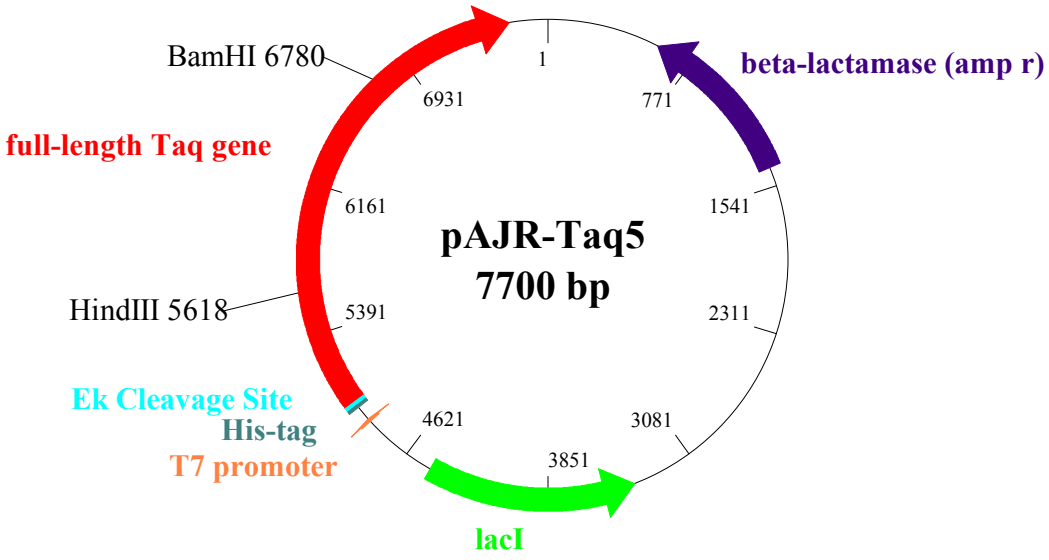


Figure 3.1. Plasmid map of pAJR-Taq5. The gene for full-length Taq DNA polymerase was cloned into the pET-46 Ek/LIC *E. coli* expression vector, and its position and orientation within the plasmid are represented by the red arrow. Overexpression yields an enterokinase cleavable His-tagged Taq fusion protein. Various other coding regions are also indicated. Beta-lactamase confers ampicillin resistance. The lacI gene codes for the lac operon repressor, which inhibits expression of the target gene until induction with isopropyl β-D-1-thiogalactopyranoside (IPTG). Unique restriction sites, BamHI and HindIII are specified because pAJR-Taq5 was digested with these enzymes to confirm its identity and size.

3.1.1.2.2 Expression and Purification

Expression and purification of His-Taq were performed using protocols previously published for untagged Taq [81] with the following modifications. *E. coli* BL21(DE3) pLysS cells (Novagen, Madison, WI) transformed with pAJR-Taq5 were grown in the presence of 200 μg/ml carbenicillin and 34 μg/ml chloramphenicol. His-Taq expression was induced with 5 mM IPTG when the absorbance at 550 nm reached 1 – 1.5 ODU. The lysis buffer, in which the cell pellets were resuspended, was comprised of 50 mM Tris-Cl, 10 mM $MgCl_2$, 50 mM glucose, 250 mM KCl, 16 mM $(NH_4)_2SO_4$ (pH 7.9). The purification procedure detailed in reference [81] was followed exactly, except the final Bio-Rex 70, pH 8.8 ion-exchange column was replaced with affinity purification of the His-Taq fusion protein over a column packed with Ni^{2+}-charged Profinity IMAC (immobilized metal affinity chromatography) resin (Bio-Rad, Hercules, CA).

The Ni^{2+}-charged IMAC column was pre-equilibrated with NiTAG-1 buffer (20 mM Tris-Cl, 22 mM $(NH_4)_2SO_4$, 10% glycerol, 1 mM imidazole, 400 mM NaCl, pH 7.9). Eluate from the immediately preceding heparin-Sepharose column was dialyzed against NiTAG-1 buffer and then loaded over the Ni^{2+}-affinity column. The column was washed with 3-4 column volumes of NiTAG-1 buffer, and purified His-Taq was eluted with a 1 – 100 mM imidazole gradient.

3.1.2 DNA Molecules

The shortest DNA used for SAXS measurements of the Taq + DNA complex was a matched 13/20mer primer-template pair with the sequence:

```
5′-TCGCAGCCGTCCA-3′
3′-AGCGTCGGCAGGTTCCCAAA-5′
```

as used previously for Taq polymerase DNA binding studies [81].

The majority of the DNA duplexes examined in isolation and in complex with Klentaq and Klenow were a matched primed-template (pt) 63/70mer and a matched blunt-ended double-stranded (ds) 63/63mer. The pt-63/70mer has an additional 7 nucleotide single-stranded overhang at the 5' end of the template strand, but the duplex sequence is identical to that of the 63/63mer:

```
5′-TACGCAGCGTACATGCTCGTGACTGGGATAACCGTGCCGTTTGCCGACTTTCGCAGCCGTCCA-3′
3′-ATGCGTCGCATGTACGAGCACTGACCCTATTGGCACGGCAAACGGCTGAAAGCGTCGGCAGGTTCCCAAA-5′
```

A fluorescently labeled DNA construct identical to the pt-63/70mer was previously used by LiCata and colleagues to study salt dependence and thermodynamics of DNA binding by Klentaq and Klenow [81, 94, 95]. However, the DNA constructs examined in these global conformation solution studies were not fluorescently tagged.

All single-stranded oligonucleotides were purchased from Integrated DNA Technologies (IDT, Coralville, IA), dissolved in STE buffer (10 mM Tris-HCl, 50 mM NaCl, 1 mM EDTA, pH 8.0), and annealed using the suggested protocol from IDT.

3.1.3 Buffers

As shown in Table 3.1, three buffers, differing in salt concentration and glycerol addition, were used in the analytical ultracentrifugation and small angle X-ray scattering experiments. The buffering agent and salt composition of these buffers were chosen based on the previous DNA binding studies of *E. coli* and *T. aquaticus* Type I DNA polymerases [81, 94, 95]. Glycerol was added to reduce radiation damage due to X-ray exposure of the protein and DNA samples during scattering experiments [96].

Table 3.1. Buffer acronyms and compositions.

Buffer Acronym	Composition
TKM-125	10 mM Tris, 125 mM KCl, 5 mM $MgCl_2$ (pH 7.9)
TKM-75	10 mM Tris, 75 mM KCl, 5 mM $MgCl_2$ (pH 7.9)
TKM-75+G	10 mM Tris, 75 mM KCl, 5 mM $MgCl_2$, 10% glycerol (pH 7.9)

3.2 Analytical Ultracentrifugation

3.2.1 Theory

Analytical ultracentrifugation (AU) is a powerful method that can be used to determine hydrodynamic and molecular properties of macromolecules in solution. In analytical ultracentrifugation, a solution of uniformly distributed macromolecules contained within a sector-shaped cell is accelerated at a particular angular velocity (ω). The resulting centrifugal field causes the macromolecules to migrate along the radial dimension away from the meniscus towards the bottom of the cell. The AU is equipped with an absorption optical system, which can radially scan the cell to monitor the distribution of macromolecules at any particular time (Figure 3.2).

The behavior of a macromolecule in solution exposed to a centrifugal field can then be used to obtain information about the molecular properties of the particle. Sedimentation velocity and sedimentation equilibrium are the two major types of experiments performed in the AU, and these techniques can be used to obtain different molecular information [97, 98].

3.2.1.1 Sedimentation Velocity

Sedimentation velocity experiments are most often used to determine a macromolecule's sedimentation coefficient, which is a hydrodynamic property of the particle. In a sedimentation velocity experiment, the rotor is accelerated with an angular velocity that is sufficiently fast, such that sedimentation dominates diffusion. Three forces act upon the sedimenting particle in the analytical ultracentrifuge: 1) a centrifugal force, produced by the spinning rotor, 2) a buoyant force, which is equal and opposite to the force required to displace a certain solvent mass, and 3) a frictional force, which results as the particle resists its flow through the solvent (Figure 3.3) [98].

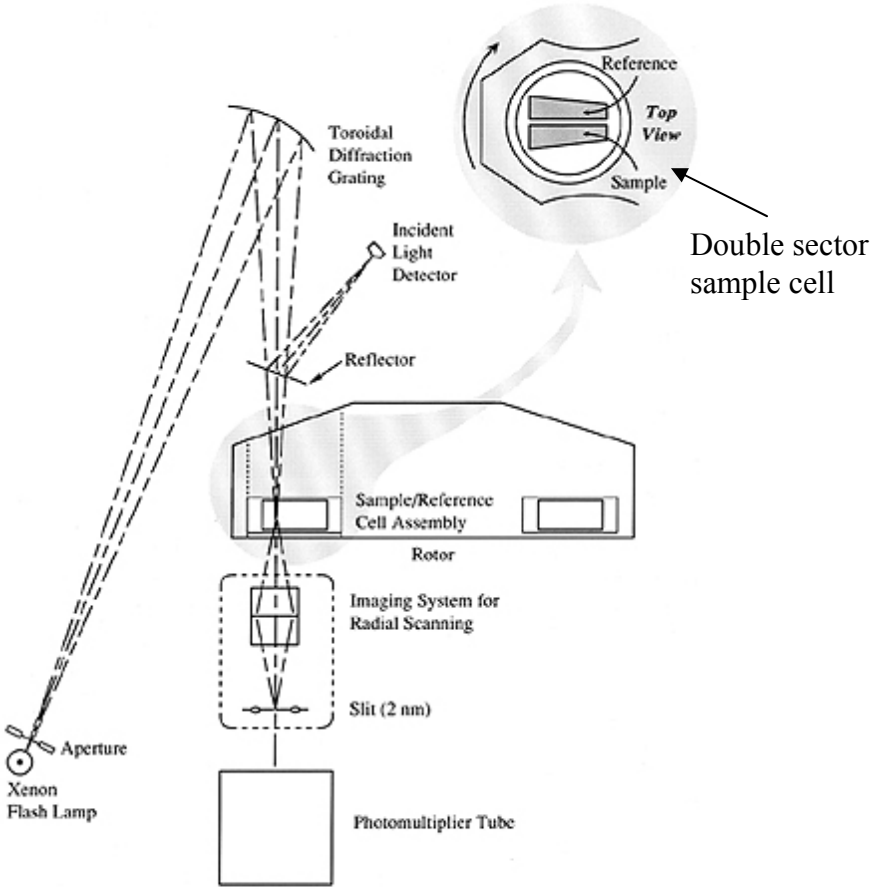


Figure 3.2. Analytical ultracentrifuge optical system. The Beckman Optima XL-A analytical ultracentrifuge is equipped with an optical system that can monitor the particle distribution within the cell by scanning the absorbance along the radial dimension of the double sector sample cell. The absorbance of the reference sector is subtracted from that of the sample sector for each absorbance measurement. (This figure was adapted from Ralston, G., Introduction to Analytical Ultracentrifugation, Beckman Instruments, Inc.)

$F_c = \omega^2 rm$ Equation 3.1
$F_b = -\omega^2 rm_o$ Equation 3.2
$F_d = -fv$ Equation 3.3

ω = angular velocity
r = distance from the center of rotation
m = particle mass
m_o = mass of displaced solution = $m\bar{\nu}\rho$
$\bar{\nu}$ = partial specific volume of the particle
ρ = solvent density
f = frictional coefficient
v = velocity of sedimenting particle

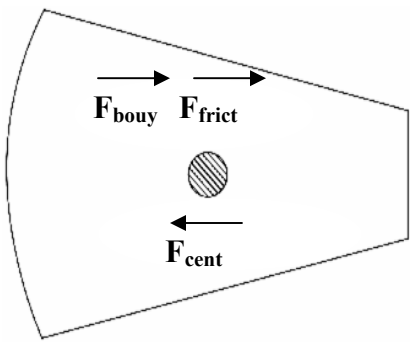


Figure 3.3. Forces experienced by a macromolecule during analytical ultracentrifugation.

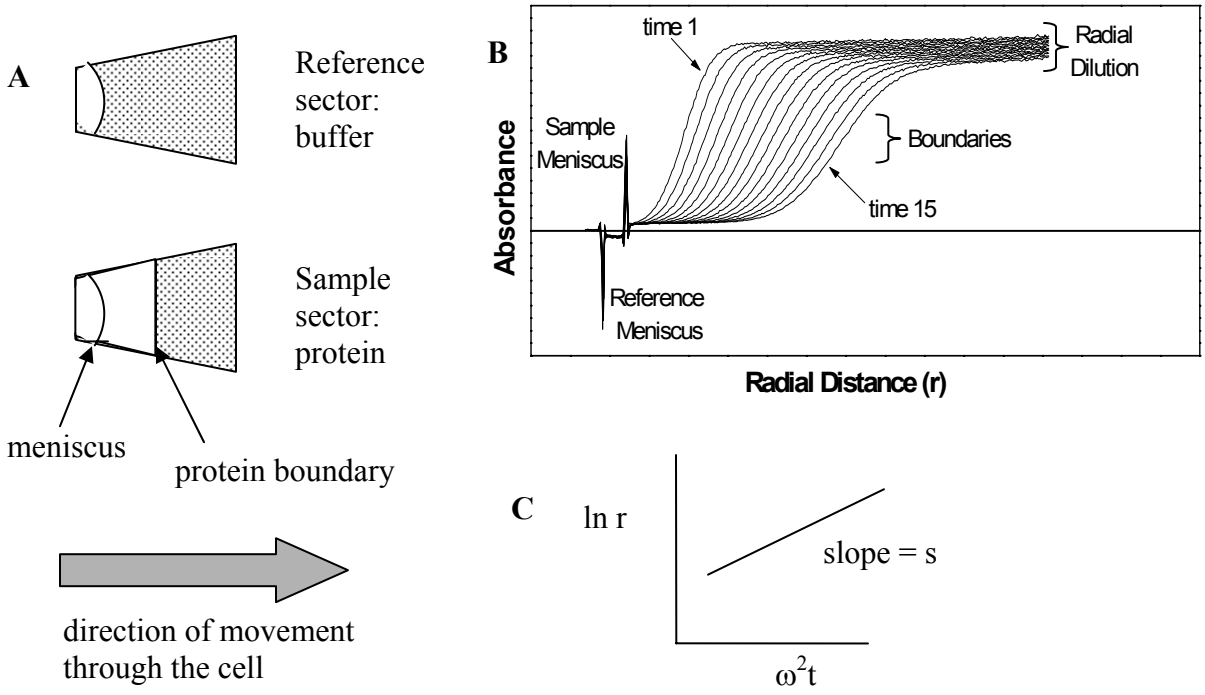
When the sedimenting particle attains a constant velocity, these three forces balance out to yield an equation in which the sedimentation coefficient (s) is related to both experimental and molecular parameters (see Figures 3.3 and 3.4 for variable definitions):

$$s = \frac{v}{\omega^2 r} = \frac{M(1-\bar{\nu}\rho)}{N_A f} \qquad \text{Equation 3.4}$$

The sedimentation coefficient is defined as the velocity (v) of the sedimenting particle divided by the strength of the centrifugal field (ω^2r), and it has the units of seconds or Svedbergs. One Svedberg is equal to 10^{-13} seconds, which is often the magnitude of the sedimentation coefficient. Experimentally, the particle velocity is equal to the radial movement of the boundary position with respect to time (dr_b/dt), and the boundary is defined as the transition midpoint of the radial absorbance profile. Replacing v with dr_b/dt in equation 3.4 and integrating over time yields Equation 3.5.

$$\ln\left(\frac{r_b(t)}{r_b(t_o)}\right) = \omega^2 s(t - t_o) \qquad \text{Equation 3.5}$$

A plot of ln $[r_b(t)/r_b(t_o)]$ versus $\omega^2 t$ has a slope equal to s (see Figure 3.4 for an overview of sedimentation velocity experiments) [97, 98].



D Equations:

$$s = \frac{M(1-\bar{v}\rho)}{N_A f} = \frac{v}{\omega^2 r} \Rightarrow v = \frac{dr}{dt} \Rightarrow s = \frac{d\ln r}{\omega^2 dt}$$

s = sedimentation coefficient
M = molecular weight
ω = angular velocity of the rotor (rps)
dr = change in radial position
r = radial distance from the center of rotation
dt = change in time

$\bar{v}$ = partial specific volume
ρ = solvent density
N_A = Avogadro's Number
v = particle/boundary velocity

Figure 3.4. Overview of sedimentation velocity. Movement of the protein in the sector cell *(A)* is followed spectrophotometrically versus time *(B)*, then plotted *(C)* so that the experimental parameters can be converted into an s value *(D)*.

According to Equation 3.4 the sedimentation coefficient is influenced by density (ρ), partial specific volume ($\bar{\nu}$), and the frictional coefficient (f), all of which depend on the temperature and solvent composition. Using Equation 3.6, the measured s value in buffer at the experimental temperature ($s_{T,b}$) is corrected to an $s_{20,w}$, which represents the sedimentation coefficient of the macromolecule in water at 20°C.

$$s_{20,w} = s_{T,b} \frac{(1-\bar{\nu}\rho)_{20,w}}{(1-\bar{\nu}\rho)_{T,b}} \left(\frac{\eta_{T,b}}{\eta_{20,w}} \right) \quad \text{Equation 3.6}$$

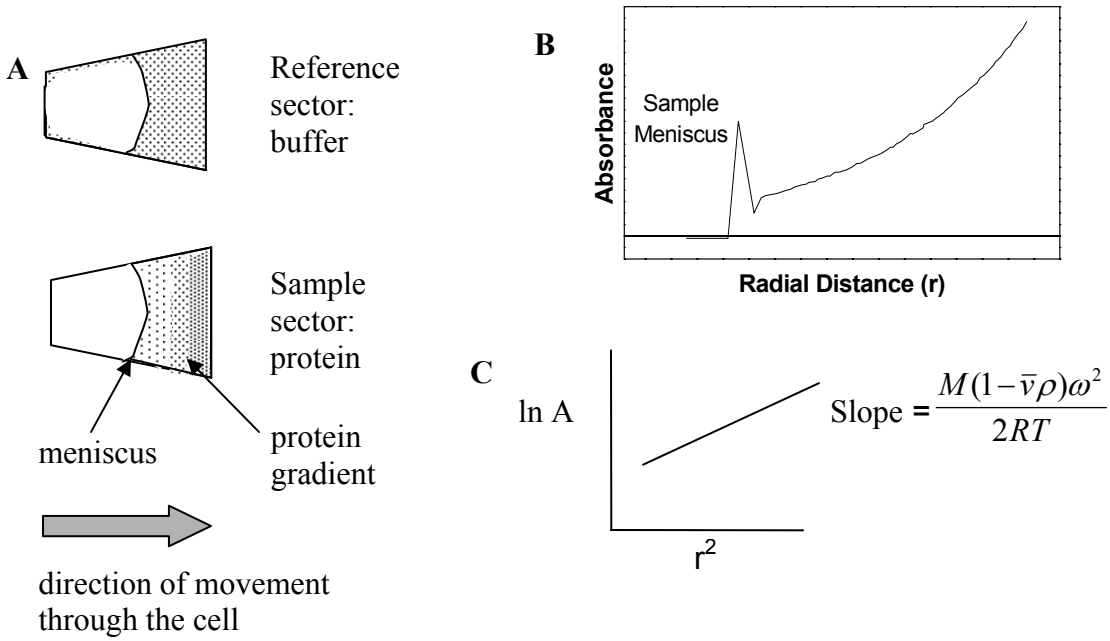
The subscripts T,b and 20,w are, respectively, used to denote parameters measured in buffer at the experimental temperature (T,b) and those measured in water at 20°C (20,w). This correction permits direct comparison of $s_{20,w}$'s measured under different experimental conditions. Use of Equation 3.6 requires accurate measurements of the solvent density and viscosity at the experimental temperature. The partial specific volume of the macromolecule dissolved in the experimental buffer should also be measured at 20°C for a more accurate $s_{20,w}$ correction. For proteins, calculation of the partial specific volume from the amino acid sequence has been well correlated with measured values, and the average $\bar{\nu}$ for proteins is about 0.73 – 0.74 cm^3/g [97-99].

3.2.1.2 Sedimentation Equilibrium

In sedimentation equilibrium experiments, the angular velocity is reduced to a speed in which the forces of diffusion are strong enough that the macromolecule does not fully sediment to the bottom of the cell, but becomes distributed in a stable concentration gradient along the radial dimension of the cell. Equilibrium experiments are useful in the determination of molecular weight, homogeneity, aggregation states, and association constants. For a single, ideal species in solution, when the equilibrium between diffusion and sedimentation is reached there is

an exponential relationship between the protein concentration (monitored by absorbance) and the square of the radial distance from the center of rotation, which is described by Equation 3.7 (see Figure 3.5 for variable definitions).

$$A_r = A_0^{\left[HM\left(x^2 - x_0^2\right)\right]} + E \qquad \text{Equation 3.7}$$



D Nonlinear Fitting Equation: $A_r = A_o^{[HM(x^2 - x_o{}^2)]} + E$

A_r = absorbance at radius x
A_o = absorbance at reference radius x_o
$H = (1-\rho)\omega^2 / (2RT)$
E = baseline offset
$\bar{v}$ = partial specific volume
ρ = density
ω = angular velocity of rotor
R = gas constant
T = temperature
M= molecular weight

Figure 3.5. Sedimentation equilibrium overview. The protein becomes distributed in a stable concentration gradient within the sector cell *(A)*, which can then be spectrophotometrically detected while spinning *(B)* and analyzed either linearly *(C)* or non-linearly *(D)*.

The molecular weight can be determined by fitting Equation 3.7 to a plot of the absorbance as a function of the square distance from the center of rotation (see Figure 3.5 for an overview of sedimentation equilibrium experiments) [97, 98].

3.2.2 Experimental Procedures

3.2.2.1 Sedimentation Coefficient Measurements

Sedimentation velocity experiments were performed in a Beckman Optima XL-A analytical ultracentrifuge. Sedimentation coefficients were measured at 20°C in TKM-125 buffer (see Table 3.1). The reference and sample sectors of Epon charcoal-filled double-sector cells were loaded with 425 μl of buffer and 400 μl of protein solution, respectively. All velocity runs were performed at 38,000 rpm in an An-60 Ti rotor for 3.5 hours. The absorbance was monitored at 280 nm. Twenty absorbance scans with a 0.004 cm step size were recorded at 10-minute intervals. Svedberg constants were determined from fits of the data to single ideal species using the program Svedberg [100, 101]. All data were well fit by a single ideal species model, across a range of concentrations. Taq concentrations examined ranged from 0.25 to 1.0 mg/ml. Klentaq concentrations examined ranged from 0.3 to 1.23 mg/ml. Klenow concentrations examined ranged from 0.36 to 1.25 mg/ml. Pol I was examined at 0.2 mg/ml only. All s values reported herein have been converted to $s_{20,w}$ values using measured solvent densities and viscosities.

3.2.2.2 Sedimentation Equilibrium Experiments and Partial Specific Volume Determination

The partial specific volumes of the polymerases were measured using the method of Edelstein and Schachman [102]. Sedimentation equilibrium experiments were performed either in H_2O/TKM-125 buffer or 96% D_2O/TKM-125 buffer (see Table 3.1). Experiments were performed in the Beckman Optima XL-A using the same rotor and cells used in the

sedimentation velocity runs. The double sector cells were loaded with 125 µl of buffer and 110 µl of protein solution in matching buffer. Equilibrium runs were performed at 9500 rpm for Pol I and Taq and 11,000 rpm for Klenow and Klentaq. All runs were carried out at 20°C for ~24 hours or until equilibrium was reached. The absorbance was monitored at 280 nm, and the initial absorbance of each protein solution was between 0.1 and 0.6. Equilibrium data were analyzed using the Origin Equilibrium analysis package provided with the instrument. The partial specific volume of each polymerase in TKM-125 buffer at 20°C was also calculated from the amino acid sequence using the computer program SEDNTERP (freeware, archived at www.bbri.org/rasmb/rasmb.html).

3.3 Density and Viscosity Measurements

TKM-125 buffer density was measured at 20°C using an Anton-Paar DMA 58 Digital Densitometer. A calibrated Cannon-Manning Semimicro kinematic viscometer was used to measure the buffer viscosity. The measured efflux time was converted to viscosity using the equation: viscosity = $t*V*\rho$, where t = efflux time, V = viscometer constant (in mm^2s^2) at temperature T (as supplied by the manufacturer), and ρ = buffer density.

3.4 Hydrodynamic Structure-based Calculations

The program HYDROPRO (version 5.a) [103] was used to calculate hydrodynamic ($s_{20,w}$) properties from the atomic coordinates in the Protein Data Bank (PDB) files of Taq (1TAQ = elongated Taq structure [30]; 1CMW = compact Taq structure [104]), Klentaq (1KTQ) [33], and Klenow (1KFD) [34]. The PDB file, molecular weight, solution density and viscosity, protein partial specific volume, temperature, radius of atomic elements (AER), and sigma values are the HYDROPRO input parameters [103]. Hydration of the protein is one of the components contributing to the AER parameter in HYDROPRO. Atomic coordinates for all heteroatoms

were deleted from each PDB coordinate file. To ensure that calculated differences between the two conformations of Taq polymerase were not due to different numbers of atoms missing in the two different structures, files were also generated that contained equal numbers of atoms: atomic coordinates missing from 1TAQ were deleted from 1CMW and vice versa. These deletions did not significantly alter the calculated values. Hydrodynamic and other solution properties were computed for each structure file using an AER value of 3 to construct the primary hydrodynamic particle, and six sigma values in a range selected to vary the number of beads in the model from ~200 to ~2000. These values are in the parameter ranges suggested by Garcia de la Torre and associates [103].

3.5 Small/Wide Angle X-ray Scattering

3.5.1 Theory

Like analytical ultracentrifugation small and wide angle X-ray scattering (SAXS and WAXS) provide structural information about macromolecules in solution. The biophysical techniques that yield the highest resolution macromolecular structures are X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. While small angle X-ray scattering yields rather low-resolution envelope-like structural models, it can provide structural information that cannot be obtained by either of the other two methods. For example, many biological macromolecules are extremely difficult to crystallize and in such cases solution scattering methods can provide structural models that aid in understanding the functional aspects of these macromolecules. Another advantage of X-ray scattering is that unlike crystallography, it allows structural determination to be performed in an aqueous solution. The buffer conditions may then be quickly and easily altered to study conformational changes that may be dependent upon ligand binding, temperature, or solution conditions such as salt type and concentration, pH, or osmotic

pressure. Structural determination by NMR spectroscopy is also performed in a solution environment, in which similar parameters may be adjusted, but effective use of this technique is limited by the size of the macromolecule. Solution structures of large macromolecules, especially those over 30 kDa, such as the Type I DNA polymerases, are very difficult to solve by NMR spectroscopy [105-107]. A collective approach that utilizes a combination of these biophysical techniques to assess a macromolecular structure often yields the most powerful results.

Efficient implementation of X-ray scattering to study dilute, aqueous solutions of bio-macromolecules requires the use of powerful synchrotron sources, which provide very high flux X-rays. These X-rays are directed through a monochromator for wavelength selection, and then the monochromatic X-rays penetrate the sample where they are scattered by the electrons of the atoms that comprise the sample. Scattered X-ray intensities are measured by the detector as a function of the scattering vector (Figure 3.6). The scattering vector, q, also known as the momentum transfer vector, is defined as $4\pi\sin\theta/\lambda$; where λ is the wavelength and 2θ is the scattering angle. According to this equation, the unit of q is the inverse wavelength ($Å^{-1}$). The scattering vector represents reciprocal space and is inversely proportional to the real space distance (d) between scattering pairs (i.e. electrons) according to the equation, $q=2\pi/d$. Thus, the one-dimensional scattering curve contains structural information concerning macromolecular shape and size, which can be extracted from the X-ray scattering pattern using an assortment of methodologies detailed in section 3.5.2.4.

3.5.2 Experimental Procedures

X-ray scattering experiments were conducted on synchrotron beam lines 1-4 and 4-2 at the Stanford Synchrotron Radiation Research Laboratory (SSRL). Preliminary data were also

collected on beam lines 1) D11A-SAXS at the Laboratorio Nacional de Luz Sincrotron (LNLS) in Campinas, Brazil, 2) 15A at the Photon Factory, Tsukuba, Japan, and 3) Bio-CAT 18-ID at the Advanced Photon Source (APS) in Argonne, IL. All data shown in the manuscript are from the SSRL beamlines but were measured during separate data collection trips (herein denoted SSRL-Dec02, SSRL-Jan03, SSRL-Jul06, and SSRL-Jun07) with different experimental conditions and beamline parameters (summarized in Table 3.2).

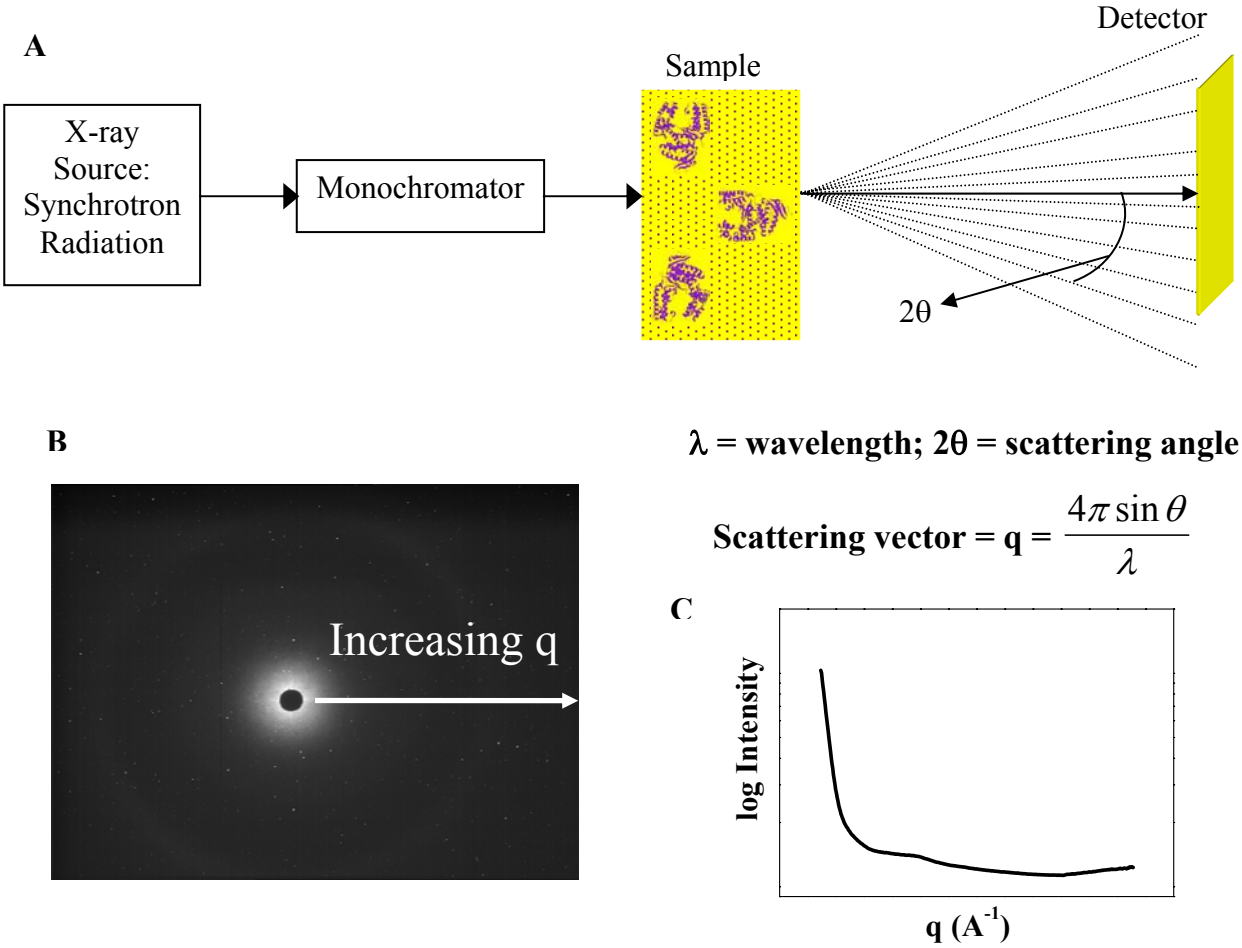


Figure 3.6. X-ray scattering overview. Simplified schematic of a synchrotron X-ray scattering beamline *(A)*. The sample is a dilute solution of randomly oriented particles. Detector distance relative to the sample is inversely proportional to the scattering angle and thus determines the range of scattering angles monitored. Scattered X-ray intensities measured by a 2-dimensional detector *(B)* are radially averaged to yield the X-ray scattering curve *(C)*. This raw scattering pattern must be subjected to additional processing, such as averaging of individual frames and buffer subtraction, prior to size and shape analysis of the scattering particle.

Table 3.2. Summary of SAXS/WAXS data sets.

SSRL Data Collection Trip	Beamline	X-ray λ (Å); Energy (keV)	Sample to Detector Distance (m)	q range ($Å^{-1}$)[b]	Samples Measured	Buffer[c]
SSRL-Dec02	4-2	1.385; 9	1.28 2	0.018 – 0.38 0.012 – 0.24	All 4 apo DNA polymerases, Taq + ddATP, Taq + 13/20mer	TKM-125
SSRL-Jan03	1-4	1.488; 8.3	0.38[a]	0.023 – 0.2	All 4 apo DNA polymerases	TKM-125
SSRL-Jul06	4-2	1.381; 9	0.5	0.04 – 1.1	Apo His-Taq	TKM-75
SSRL-Jun07	4-2	1.127; 11	2.5 1	0.008 – 0.3 0.03 – 0.77	Lysozyme, Apo Klentaq and Klenow, Isolated 63/70mer and 63/63mer, Binary Complexes listed in Table 3.3	TKM-75+G

[a] Optimal sample to detector distance for this beamline.
[b] $q \leq 0.3Å^{-1}$ – SAXS; $q \geq 0.3Å^{-1}$ – WAXS
[c] See Table 3.1 for buffer compositions.

3.5.2.1 Preparation of Macromolecules and Complexes for X-ray Scattering Experiments

3.5.2.1.1 Isolated Polymerases and DNA

Polymerases and DNA constructs were dialyzed against the buffers as indicated in Table 3.2. Following dialysis, each apo polymerase was concentrated to very near its solubility limit, typically between 5 – 10 mg/ml, using Amicon® centrifugal filter devices with a 30 kDa molecular weight cutoff (Millipore, Billerica, MA). However, freshly purified full-length Taq polymerase would often begin to precipitate at concentrations above 1.5 mg/ml, and in these cases, further concentration was not attempted. Any insoluble material was removed from the protein samples by microcentrifugation at 14000 rpm for 30 minutes. Occasional additional centrifugation cycles were carried out when the resulting pellet was loosely adhered to the microfuge tube wall. Using molar extinction coefficients (ε_{280}) of 87650 M^{-1} cm^{-1} for full-length Pol I [108], 112690 M^{-1} cm^{-1} for full-length Taq and His-Taq, 70360 M^{-1} cm^{-1} for Klentaq and 58790 M^{-1} cm^{-1} for Klenow, polymerase concentrations were measured by UV spectroscopy at 280 nm on a Varian Cary 100 UV-Vis spectrophotometer. The protein extinction coefficients were obtained from cited literature or calculated from the respective amino acid sequences using the *ProtParam* online tool (http://ca.expasy.org/tools/protparam.html) [109].

DNA concentrations were determined by UV spectroscopy at 260 nm (1 μg/ml dsDNA = A_{260} ~0.02 ODU) [110]. DNA stock concentrations were typically ≥ 800 μM. The purity and stoichiometric strand match of each DNA stock was monitored by non-denaturing polyacrylamide gel electrophoresis, where each construct migrated as a single band (data not shown).

All macromolecules were stored at ~4°C during transport to the synchrotron. Prior to obtaining X-ray scattering data at SSRL, the protein samples were again centrifuged to remove

any precipitates that may have formed during transport and their concentrations were re-measured by UV spectroscopy using a Varian Cary 50 spectrophotometer. The pre- and post-transport stock concentrations were typically very similar. Scattering curves were typically collected at 3 – 5 concentrations as specified. Each concentration series was generally prepared by serial dilution.

3.5.2.1.2 Binary Polymerase + DNA Complexes

Acronyms used for the binary complexes are denoted in Table 3.3.

Table 3.3. Binary Complexes Examined.

Binary Complex	Components
ptKLN	Klenow + pt-63/70mer
dsKLN	Klenow + ds-63/63mer
ptKTQ	Klentaq + pt-63/70mer
dsKTQ	Klentaq + ds-63/63mer

At least 15 minutes prior to data collection, each binary complex was prepared by mixing buffer with equimolar amounts of protein and DNA from the stock solutions in 1.5 ml microfuge tubes to yield an 80 μM solution. Polymerase + DNA complexes were assayed at 3 – 5 concentrations between 10 - 80 μM to check for the presence of aggregation or interparticle interference. Subsequent concentrations were prepared by serial dilution. With one exception, the concentrations assayed in these X-ray scattering experiments were more than 500-fold over the Kd's (dissociation constants) for Klentaq binding to 63/70mer [94] and 63/63mer (Wowor and LiCata, unpublished); and more than 1000 times higher than the Kd's of Klenow binding to

63/70mer [81, 95] and 63/63mer (Wowor and LiCata, unpublished). For the one exception, 10 μM ptKTQ, the concentration was approximately 250 times greater than the measured Kd [94].

3.5.2.1.3 Lysozyme

Lysozyme was used as a calibration standard to correlate molecular weight with forward scattering intensity, also known as I(0). An approximately 30 mg/ml lysozyme stock solution was prepared by directly dissolving the lyophilized powder in TKM-75+G buffer, which was the same buffer used for the other SSRL-Jun07 samples. Any insoluble material was then removed by centrifuging the stock at maximum speed in a microcentrifuge for 30 minutes. The UV absorbance of the supernatant was measured at 280 nm in a Varian Cary 50 spectrophotometer and converted to concentration using the extinction coefficient of 2.64 ml mg^{-1} cm^{-1} [111]. Four concentrations of lysozyme ranging from 2.5 - 15 mg/ml were prepared by serial dilution and measured by SAXS. Only the molecular weights of the SSRL-Jun07 samples were checked by calibration against I(0).

3.5.2.2 Individual Data Set Collection and Data Reduction Details

Over the past 4 – 5 years since we began our synchrotron X-ray scattering studies, the technique has undergone major developmental changes to improve its ease of use by a broad range of researchers. Enhanced automation of data collection, data treatment, and preliminary analysis to deal with the large volumes of data obtained during a single data collection trip is steadily transitioning X-ray scattering into a high throughput biophysical technique. The most efficient experimental protocol (that of SSRL-Jun07) that results in little to no X-ray induced sample damage and rapid data treatment is presented in detail, followed by important experimental parameters from the other data collection trips denoted in Table 3.2.

3.5.2.2.1 SSRL-Jun07 Data Collection, Treatment, and Reduction

To minimize the possibility of X-ray induced sample damage, a flow cell was used for these scattering measurements. Flow cell sample chambers significantly reduce radiation damage associated with high flux third generation synchrotron sources. High flux beamlines facilitate collection of high quality wide angle scattering data by significantly improving the signal to noise ratio of the scattering curve, and reduce exposure time, which aids in buffer subtraction problems associated with beam movement (instability) between background and sample measurements. However, the price is a higher chance of radiation damage, which can be significantly lessened by oscillating or flowing sample through the beam during X-ray exposure.

The flow cell system, designed by the SSRL BL 4-2 beamline scientists, consisted of a 1.5 mm diameter quartz capillary (Charles Supper Company, Natick, MA) housed within a temperature controlled unit and mounted within the beam path. Tubing connected the input side of the capillary to a robotically controlled needle and the output side to a Hamilton MICROLAB dual syringe dispenser. Both the Hamilton sample dispenser and the automated needle were integrated with the distributed control system (DCS) and the data acquisition software through the Blu-ICE graphical user interface (GUI), and controlled by a customized command script. The DCS and Blu-ICE GUI were originally designed and developed for crystallography beamlines by the Structural Molecular Biology (SMB) group at SSRL, and then subsequently adapted to control X-ray scattering data acquisition (detailed information concerning this system and software can be found at (http://smb.slac.stanford.edu/public/research/developments/blu-ice/intro.html). A PCR block was used to hold multiple sample tubes and the cleaning solution, which was a 10% bleach solution. The robotically controlled needle moved to specified locations within the PCR block to load or discard sample or cleaning solution.

Sample loading/unloading and data collection were linked in Blu-ICE and robotically controlled by the user generated script. All buffer and sample scattering measurements were performed in a similar manner. A volume of 30 μl was loaded into the flow cell capillary. During X-ray exposure, 24 μls were continuously oscillated through the beam at a flow rate of 1 μl/s. Following each measurement, the sample was dispensed into a fresh tube for further analysis if needed. Scattering profiles were collected for each buffer-sample pair without re-entering the hutch. The flow cell was cleaned after each buffer-sample pair by oscillating a 10% bleach solution back and forth through the capillary and then flowing buffer through the system.

As noted in Table 3.2, both small and wide angle X-ray scattering curves were measured for each sample, except lysozyme, for which only SAXS data were collected. The beam flux was ~ 10^{11} photons/s. Scattered X-rays were detected with a position-sensitive MarCCD 165 detector operated in "dezingered" mode (which automatically subtracts intensity spikes caused periodically by cosmic rays randomly hitting the detector), and slow mode with a resolution of 512 x 512 pixels. Dark counts were automatically subtracted from the CCD (charge-couple device) image for each collection frame. Beamline computer programs, Blu-ICE and MarParse, were used, respectively, for data acquisition and preliminary data treatment.

In addition to the user customized flow cell control (FCC) script, other Blu-ICE user input parameters included the number of frames and exposure time per frame. While continuously oscillating through the X-ray beam at a rate of 1 μl/s, each buffer/sample was exposed for 10 seconds per frame, and 10 frames were collected. This data collection strategy was used to check for possible X-ray induced sample damage, which typically manifests itself as a time dependent variation in the scattering curve (see Figure 3.7). Under the conditions of these experiments, X-ray induced sample damage was not observed, i.e. consecutive scattering curve

frames did not reveal patterned variation that was dependent upon the X-ray exposure time. However, random aberrant frames were observed periodically and could often be attributed to a bubble passing through the beam (see Figure 3.8). When excessive, random deviation between subsequent frames was noticed, replacing the cleaning solution typically solved this problem.

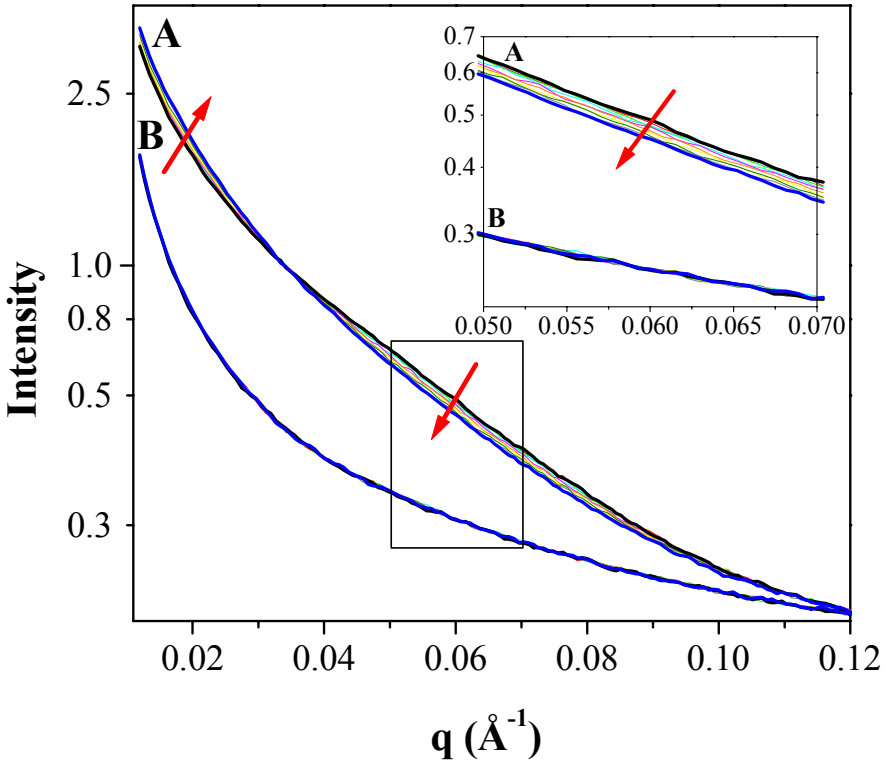


Figure 3.7. Detection of time-dependent effects of X-ray induced aggregation on the scattering curve. *Series A:* aggregating macromolecule. *Series B:* No aggregation. Series A and B each contain 10 successive scattering frames collected during 30 second X-ray exposures from different macromolecules. Curves corresponding to frames 1 and 10 are bolded and colored black and blue, respectively. Other colors represent intermediate frames. For the aggregating sample *(A)*, at very low q values (left arrow) the scattering pattern increases in intensity as a function of increasing X-ray exposure time. As q becomes larger, the successive frames cross one another and then the pattern reverses such that scattering intensity decreases with increasing exposure (right arrow). At even higher q values, the curves all merge and no longer exhibit X-ray induced variation. The q values where the successive scattering curves cross and merge are dependent upon the specific macromolecule and exposure time. For the non-aggregating sample *(B)*, the 10 successive data frames all overlap. *Inset:* Magnification of the boxed region.

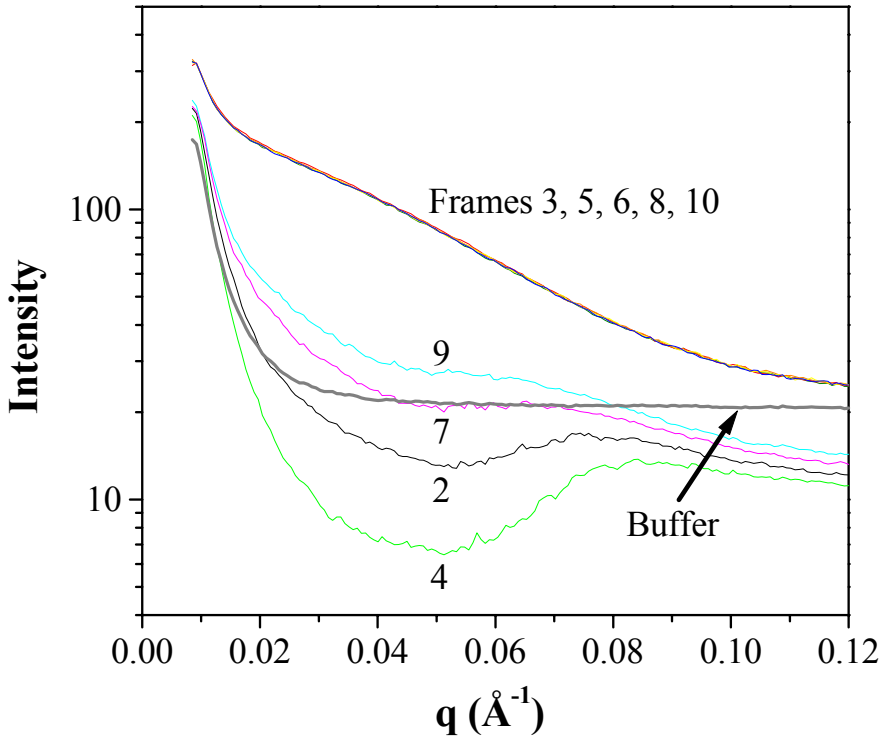


Figure 3.8. Example of random scattering curve deviation. Scattering curves 2 – 10 were collected during consecutive 10 seconds X-ray exposures as macromolecules flowed through the sample capillary. The bolded gray curve is the averaged buffer scattering pattern. Frames 2, 4, 7, and 9 exhibit strange patterns and have scattering intensities that dip below that of the buffer scattering curve. We attributed these anomalous scattering patterns to bubbles flowing through the X-ray beam. In this example, frames 3, 5, 6, 8, and 10 were unaffected and were averaged and used for further analysis. Typically, 2 – 5 out of 10 curves were irregular.

The log file and 2-D CCD scattering image generated by Blu-ICE for each acquisition frame were inputted into MarParse to 1) integrate and normalize the intensities from the CCD image and 2) convert the detector channels to q. To correct for small changes in beam flux and differences in sample absorption, all scattering intensities were normalized by the transmitted beam intensity integrated over the X-ray exposure time. The detector channel to q conversion was determined by measuring the standard scattering of silver behenate ($C_{21}H_{43}COOAg$) powder (provided by beamline scientists), which has a well-defined D-spacing of 58.380 Å [112-114]. For a crystalline powder, such as silver behenate, or any crystal in general, the D-spacing is the

perpendicular distance separating the individual planes of a crystalline lattice. Since such substances produce a well defined and quantized I(q) pattern, they are used to calibrate pixel positions on the CCD image to the corresponding scattering vector (q). Buffer and macromolecule SAXS/WAXS data curves were all analyzed in a similar fashion using MarParse to yield scattering intensity (with associated error) as a function of q for each frame.

Data reduction was performed using the computer program, PRIMUS [115], which is a part of the ATSAS 2.1, small-angle scattering analysis package [116]. For each buffer or sample, all of the MarParse processed scattering curve frames were graphed and visually inspected for X-ray induced scattering curve variation as a function of exposure time and/or random aberrant scattering profiles. For the SSRL-Jun07 data, when compared to subsequent scattering curves, frame 1 from all non-buffer samples was consistently shifted to slightly higher scattering intensities in both SAXS and WAXS experiments. This scattering curve variation was not indicative of X-ray damage. Since the buffer scattering data were not similarly affected, we attributed this difference to possible sample dilution as the new sample enters the capillary and automatically discarded the first frame for all non-buffer samples. All other visually overlapping scattering curve frames were averaged and then buffer subtracted using either PRIMUS or MarParse.

3.5.2.2.2 Pre-SSRL-Jun07 Data Collection, Treatment, and Reduction

SSRL-Dec02 and SSRL-Jan03 scattering curves were measured in multiple frames over a total exposure time of 15 minutes. Scattering data were collected, normalized for dark counts and transmitted intensity, and the buffer background scattering was subtracted using the resident analysis programs at each beamline. A total exposure time of 5 minutes (10, 30-second frames)

was used to measure the SSRL-Jul06 X-ray scattering patterns, and data acquisition, treatment, and reduction were very similar to that of SSRL-Jun07.

For SSRL-Dec02, the sample cell was aluminum and consisted of a 200 µl flat sided, round sample chamber with a 1.4 mm path length. The sample chamber was bounded on both sides by circular Kapton windows. SSRL-Jan03 and Jul06 scattering measurements were performed in plastic, type I (Lexan) sample cells with flat mica windows and a ~30 µl sample volume (http://www-ssrl.slac.stanford.edu/~saxs/instrumentation/more/more.htm#samplecells). All of these measurements were static, in contrast with the flow cell measurements of SSRL-Jun07.

For the SSRL-Dec02 and SSRL-Jan03 scattering data, no exposure dependent X-ray radiation damage was observed (i.e. all frames overlapped), and thus all SAXS frames were averaged for each sample. Also, the experimentally determined Rg (radius of gyration) values did not show any protein concentration dependence within error. SSRL-Jul06 scattering data, however, exhibited radiation damage with increased X-ray exposure, similar to the time dependent damage shown in Figure 3.7. Subsequent frames did not superimpose with the first frame; therefore, only the first frame of each SSRL-Jul06 data set was used for further analysis.

3.5.2.3 Data Analysis and Modeling

Structural information was extracted from the X-ray scattering curves using four major analysis methodologies. The radius of gyration (Rg) and I(0) (also known as the forward or zero-angle scattering) were determined using both 1) the Guinier approximation [117, 118] and 2) the indirect, inverse Fourier transformation algorithm implemented by the GNOM computer program [119, 120]. 3) Using the program CRYSOL [121], a simulated scattering curve was generated from a known X-ray crystal structure or a user constructed atomic coordinate model

(ACM) built from crystallographic and/or simulated atomic coordinates, and then compared with or fit to the measured scattering curve. 4) Three dimensional *ab initio* shape models composed of spherical volume elements (beads) were constructed from the experimental scattering curve using the computer programs, DAMMIN/DAMMIF [107] and/or GASBOR [122].

Some, but not all of the scattering curves exhibited a sharp spike in intensity at very low q values ($q < 0.02$ Å^{-1}), which was removed before analysis. This behavior was only present in some of the beamline configurations used, and seemed to be best explained as parasitic scattering around beam stop. This behavior could not be attributed to aggregation since 1) it was sometimes absent; 2) it was generally identical from protein to protein, including control proteins such as aspartate transcarbamylase (data not shown); and 3) it was not correlated with the protein concentration. Furthermore, the equilibrium centrifugation experiments also conducted on the apo DNA polymerases showed no appreciable protein aggregation.

3.5.2.3.1 Rg and I(0) Determination

According to the Guinier equation (see Equation 3.8), I(0), which is proportional to the molecular mass of the scattering particle (see Section 3.5.2.3.2), and Rg are calculated from the linear region of a Guinier plot, in which the natural logarithm of the intensity is graphed as a function of the scattering vector squared. The Guinier approximation is only valid at small scattering angles where $Rg*q_{max}$ is less than or equal to 1.5 [106, 123, 124].

$$\ln I(q) = \ln I(0) - \frac{Rg^2 q^2}{3} \quad \text{Equation 3.8}$$

For rigid rods or cylindrical shaped particles, in which one dimension is at least twice as long as the other two, like the DNA constructs and binary complexes investigated in these studies, the radius of gyration of the cross-section (Rg_{xs}) can be calculated using a Guinier-type approximation for the cross-section. Multiplication of the scattering intensity by q eliminates the

length information from the long dimension and yields Rg_{xs} pertaining to the scattering particle's cross-section. According to this analysis, a plot of ln[I(q)*q] versus q^2 yields a linear region with a slope equal to $-Rg_{xs}^2/2$; $Rg_{xs}*q_{max}$ should be less than or equal to 1.5 [105, 106, 125].

In GNOM analysis of a system of particles that are monodisperse in both size and shape, the maximum dimension (Dmax) of the macromolecule is altered by the user until a suitable solution is found. The inverse Fourier transform of the X-ray scattering curve computed by GNOM is the P(r) distance distribution function, which describes the distribution of distances between scattering pairs, and goes to zero at Dmax [119]. Rg and I(0) are then calculated from the P(r) function according to equations 3.9 and 3.10 [124]. GNOM versions 4.4 and 4.5 were used for our analyses.

$$I(0) = 4\pi \int_0^{D\max} P(r)dr \qquad \text{Equation 3.9}$$

$$Rg^2 = \frac{\int_0^{D\max} r^2 P(r)dr}{2\int_0^{D\max} P(r)dr} \qquad \text{Equation 3.10}$$

3.5.2.3.2 Molecular Mass Determination - Calibration against Lysozyme I(0)

Scattering intensity I(q) is directly proportional to the molecular mass (M) of the scattering particle according to equation 3.11,

$$I(q) = \frac{M \cdot \Delta z^2 \cdot c \cdot r_e^{\,2} \cdot t \cdot N_A \cdot F}{d^2} \cdot P(q) \qquad \text{Equation 3.11}$$

where c is the mass concentration, r_e is the classical radius of an electron (2.8179×10^{-13} cm) [126], t is the sample thickness, N_A is Avogadro's number, F is the incident beam flux or total energy per unit time irradiating an area of the sample, P(q) is the form factor of the scattering particle, which contains information concerning particle shape and structure, d is the sample to

detector distance, and Δz is the effective number of mole electrons per gram of the scattering particle in excess over the buffer electron density composition. To calculate Δz (see Equation 3.12), the partial specific volume of the macromolecule ($\bar{v}_{sp}$) is multiplied by the mean electron density of the buffer (ρ_0 in mol e^-/ml) and then subtracted from the number of mole electrons per gram composing the scattering particle (z_{sp}) [123].

$$\Delta z = z_{sp} - \bar{v}_{sp}\rho_0 \qquad \text{Equation 3.12}$$

At zero scattering angle, i.e. when q = 0, the form factor, P(q), becomes P(0) and is equal to 1, regardless of particle shape [123]. For a specific sample to detector configuration and sample cell, t, F, and d remain constant and r_e and N_A are also constants, thus the difference in the I(0) between two samples measured in the same sample cell with the same beamline setup is dependent solely upon variation in M, Δz, and c, irrespective of particle shape ($I(0) \propto M \cdot \Delta z^2 \cdot c$).

When compared to a protein standard (std) with known molecular weight, the molecular mass of the scattering particle (sp) is calculated according to Equation 3.13.

$$M_{sp} = (I(0)/c)_{sp} / \Delta z_{sp}^{\ 2} \cdot \frac{M_{std}}{(I(0)/c)_{std} / \Delta z_{std}^{\ 2}} \qquad \text{Equation 3.13}$$

For our experiments, we used lysozyme as the protein standard to calibrate the proportionality between molecular weight and the forward scattering intensity. To ensure that our lysozyme standard was monomeric and monodisperse, we assayed a series of concentrations and compared the experimentally determined Rg to literature values [127, 128]. In this investigation, the forward scattering intensity I(0) was normalized by concentration and then extrapolated to infinite dilution yielding $I(0)/c_{(c=0)}$, which was subsequently normalized against Δz^2 to obtain $(I(0)/c)_{c=0}/\Delta z^2$. Thus, the molecular masses in this study were determined by replacing the $(I(0)/c)/\Delta z^2$ values in Equation 3.13 with $(I(0)/c)_{c=0}/\Delta z^2$.

Molecular weight determination from SAXS data is very sensitive to the partial specific volume of the macromolecule, which is used to calculate Δz^2 (see Equations 3.12 and 3.13). The average empirically determined partial specific volume cited for proteins is 0.73 – 0.74 ml/g [97-99]. However, $\bar{\nu}$ calculations from the primary amino acid sequence and actual $\bar{\nu}$ measurements reveal that the partial specific volume is much more variable. For example, SEDNTERP calculated $\bar{\nu}$ for lysozyme is 0.7175 ml/g, and measured $\bar{\nu}$ values are between 0.712 – 0.742 ml/g [129, 130]. Although partial specific volume values calculated from primary sequence or measured should be logically more accurate, according to a study performed by Svergun and colleagues on a set of proteins with a broad range of molecular masses, a $\bar{\nu}$ of approximately 0.74 ml/g should be used for all proteins to calculate molecular mass from SAXS data [128].

Since partial specific volume and electron density are typically independent of specific protein composition, normalization of I(0) by the excess electron density contrast (Δz) is not necessary if both the scattering particle and the standard are proteins. In this scenario, Equation 3.13 will simplify to

$$M_{sp} = (I(0)/c)_{sp} \cdot \frac{M_{std}}{(I(0)/c)_{std}} \qquad \text{Equation 3.14}$$

However, compared to proteins, DNA molecules have a higher electron density and a lower partial specific volume (Table 3.4). Therefore, we have normalized the $(I(0)/c)_{c=0}$ values from the isolated DNA molecules and the polymerase + DNA complexes by Δz^2 to accurately determine the molecular weight of these scattering particles. Similar techniques have been used by other researchers to determine molecular weight or concentration of protein-DNA complexes from solution scattering data [131, 132].

Table 3.4. Physicochemical properties of proteins, DNAs, and binary complexes examined.

Macromolecule or Complex	Partial specific volume (ml/g)[a]	Electron density ($e^-/Å^3$)	z (mole e^-/g)[a]	Δz (mole e^-/g)[c]
Protein	0.74	0.435	0.535	0.113
DNA	0.52[b]	0.598	0.516	0.220
ptKTQ	0.653	0.487	0.527	0.155
dsKTQ	0.656	0.485	0.528	0.154
ptKLN	0.657	0.484	0.528	0.153
dsKLN	0.660	0.482	0.528	0.152

[a] z and $\bar{v}$ values for the binary complexes are weighted averages of the respective values for the components alone [133]
[b] DNA $\bar{v}$ from reference [132]
[c] Calculated according to Equation 3.12; ρ_o = 0.5703 mol e^-/ml for TKM-75+G buffer

3.5.2.3.3 Shape Modeling

Scattering curves used for *ab initio* shape modeling were prepared from the SSRL-Jun07 scattering data for the binary complexes, apo Klenow, and isolated DNA molecules by merging the small angle region (q range from 0.008 to 0.14 $Å^{-1}$) of the curve from the low concentration (1.2 – 2.2 mg/ml) sample with the wide angle region (q range from 0.07 – 0.76 $Å^{-1}$) from the high concentration (3 – 9 mg/ml) data. All curves were merged using the program PRIMUS [115]. Combining the scattering curves in this way, as shown in Figure 3.9, yields a final scattering curve for modeling that consists of both high quality (i.e. high signal to noise ratio) wide angle data, and small angle data that is relatively free from concentration dependent behavior. For apo Klentaq, the wide angle scattering curve was created by linear extrapolation of the non-buffer subtracted concentration series curves to infinite dilution using PRIMUS. Such extrapolation can help to eliminate the subtle effects of excluded volume on the X-ray scattering curve within the wide angle regime [115]. The extrapolated wide angle scattering curve was then merged with the low concentration small angle data for apo Klentaq using the merging procedure discussed. Similar extrapolations could not be performed for the other scattering particles because of poor quality wide angle scattering data at the lower particle concentrations.

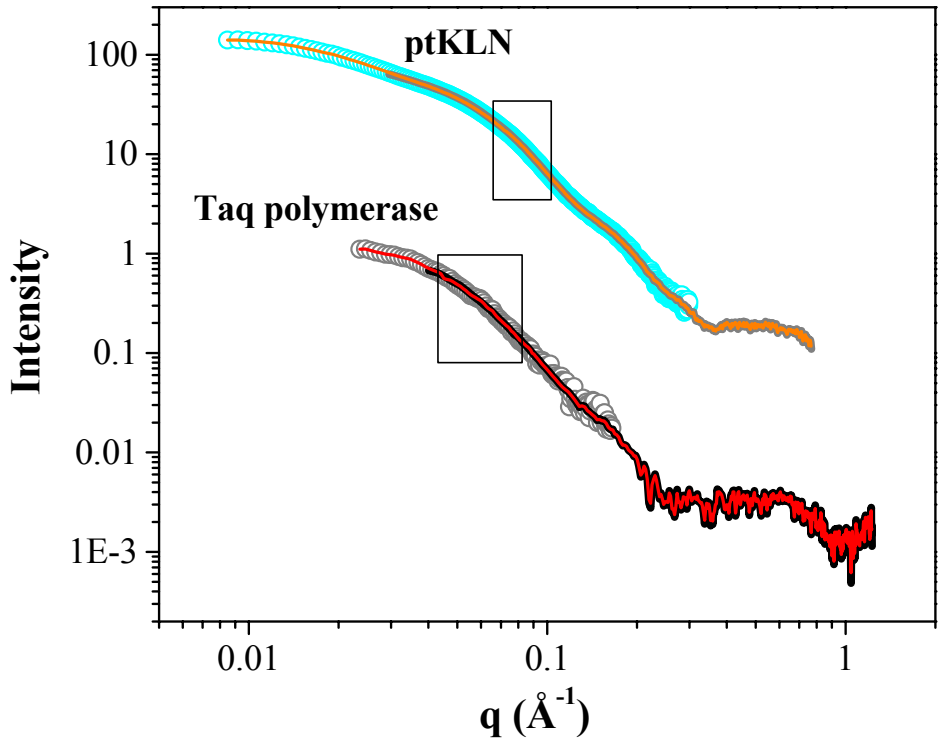


Figure 3.9. Merging of small and wide angle scattering data. This figure contains scattering curves from ptKLN and full-length Taq DNA polymerase collected over 2 different q ranges and sample concentrations. X-ray scattering curves collected from 2.2 mg/ml ptKLN (*cyan open circles*), 8.7 mg/ml ptKLN (*gray solid line*), ~5 mg/ml wild-type Taq *(gray open circles),* and ~5 mg/ml His-Taq (*black solid line)* are plotted on a log-log scale to emphasize the overlapping region within the small angle regime. While ptKLN scattering profiles were collected on the same beamline with the same buffer conditions, full-length Taq scattering data were collected on different beamlines under slightly different buffer conditions. The rectangles identify the superimposing regions (which typically contained 10 – 60 data points from each curve) used by PRIMUS [115] to combine the low and high q data for each scattering particle. All data points outside of the rectangle 1) on the left belonging to the high q curve and 2) on the right belonging to the low q curve were excluded from the merged curve. The resulting merged scattering curves for ptKLN *(orange solid line)* and Taq *(red solid line)* are also shown. (ptKLN is a much larger particle than full-length Taq; thus, the scattering profile for ptKLN was measured to much lower q values than Taq.)

The merged scattering curve for full-length Taq DNA polymerase was similarly constructed using the small angle (q range from 0.024 – 0.09 $Å^{-1}$) and wide angle (q range from 0.05 – 1.23 $Å^{-1}$) regions, respectively, from SSRL-Jan03 and Jul06 scattering curves for Taq and His-Taq. Sample concentrations for full-length Taq and His-Taq were both approximately 5

mg/ml. Although the low angle scattering data for Taq was collected from a relatively highly concentrated sample, no interparticle interference or radiation induced aggregation was observed under these experimental conditions. While Taq and His-Taq scattering curves were collected in different buffers on different beamlines, excellent superimposition of the Taq (low q regime) and His-Taq (high q regime) buffer-subtracted scattering curves suggests that the global conformation of full-length Taq DNA polymerase is not significantly altered by either the N-terminal His-tag or the buffer conditions (see Figure 3.9). Thus, it was deemed acceptable to merge the scattering curves of wild-type and His-tagged Taq for shape modeling.

Merged scattering curves were subjected to GNOM analysis, and then global conformational models of each macromolecule and macromolecular complex were created via execution of the *ab initio* bead modeling programs, DAMMIF [107] (version 0.9) for DNA constructs and binary complexes and GASBOR [122] (version 2.2i) for apo polymerases. DAMMIF is a recoded computationally faster version of its predecessor program DAMMIN. Both programs yield reconstructed bead models of equivalent quality.

Like DAMMIN, DAMMIF begins with a search volume filled with a very large number of relatively small densely packed beads arranged on a grid [107]. For the traditional globular protein, the initial search volume is typically a sphere (see Figure 3.10) with a diameter that is equivalent to the maximum dimension of the particle. However, for rod-like macromolecules, such as the isolated DNA constructs and binary complexes examined in this study, a cylindrical initial search volume (see Figure 3.10) is more suitable, where the cylinder height (i.e. Dmax), inner radius, and outer radius must be specified by the user. Since DNA is not characterized by a hollow interior, the inner radius of the cylinder was set to zero. The outer radius and cylinder height (or sphere radius if using a spherical search volume) were set to values slightly larger than

the expected particle dimensions to avoid unnatural flattening of the dummy atom model (DAM) near the edge of the search volume (see Figure 3.10).

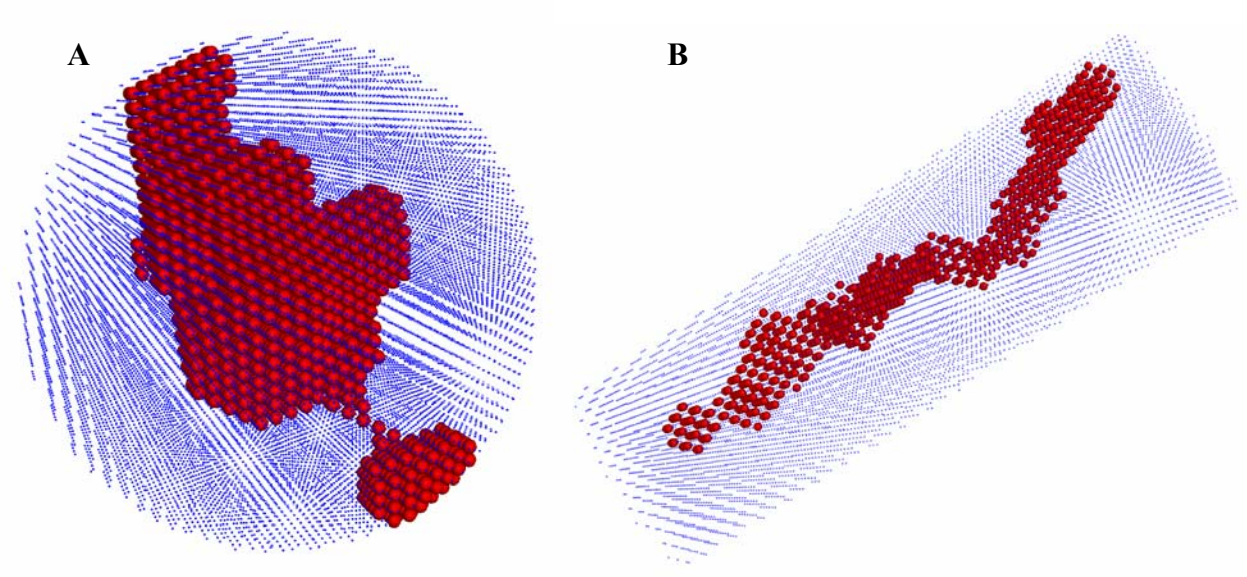


Figure 3.10. Search volume examples and edge-effects in DAMMIN/DAMMIF. Solvent beads are blue and particle beads are red. *A:* Spherical search volume, where the diameter of the sphere is equal to the maximum dimension of the particle. It is obvious in this figure that the particle shape is artificially flattened at the edges of the search volume and this is why the size of the initial search volume should be slightly larger than the expected particle dimensions. *B:* Cylindrical search volume with outer radius and cylinder height slightly larger than particle dimensions; no edge-effects are observed.

DAMMIN/DAMMIF employs a two-phase system, in which each bead belongs either to the solvent (0) or the particle (1). During a simulated annealing global minimization procedure, each bead switches one at a time between solvent and particle designations to minimize the function denoted as Equation 3.15 and yield a compact, interconnected bead model whose scattering pattern matches the experimental data [107].

$$f(X) = \chi^2 + \alpha P(x) \qquad \text{Equation 3.15}$$

In equation 3.15, the function χ^2 describes the discrepancy between the DAM and experimental scattering curves, α is a weighting factor, and P(x) is a penalty function. P(x) is composed of individual weighted penalties, the looseness penalty and the peripheral penalty, which increase when the DAM is, respectively, less compact and more elongated (i.e. less globular) [107]. Since we expect the isolated DNA molecules and binary complexes to be more rod-like than globular, the penalty weight for the peripheral penalty was set to zero. The wide angle scattering data ($q > 0.3$ Å^{-1}) was not used for DAMMIF modeling. User defined DAMMIF input parameters, aside from those previously mentioned (shape of search volume, search volume dimensions, and peripheral penalty weight), are listed in Table 3.5.

Table 3.5. DAMMIF input parameters for shape modeling of isolated DNA and binary complexes.

DAMMIF parameter[a]	Parameter Value	
	Isolated DNA	Binary Complex
Outer cylinder radius (Å)	40	60
Cylinder height (Å)	250	280
Number of dummy atoms	20000	40000
Number of harmonics	25	
Looseness penalty weight	0.005	
Maximum number of temperature steps in annealing[b]	400	
Minimum number of successes per temperature step	20	

[a] The default value was used for any parameter not noted herein.
[b] Slow mode default values were used for all other annealing parameters.

While the simulated annealing minimization technique implemented in GASBOR is very similar to that of DAMMIF, the most significant difference between these two shape modeling programs lies in the description of the scattering particle. Unlike DAMMIF, GASBOR defines the scattering particle as protein and initializes the model as a chain-like assembly of beads

(dummy residues), in which each dummy residue (DR) represents a Cα atom and the total number of DRs corresponds to the number of amino acid residues comprising the protein. The distance separating adjacent DRs is equivalent to the observed spacing between Cα atoms in proteins (~ 3.8 Å) [122]. This description of the scattering particle as a specifically spaced chain-compatible ensemble of dummy residues (DRs) precludes the use of GASBOR for modeling scattering particles composed of DNA. To fit the experimental scattering curve, a randomly selected DR is moved within a spherical search volume (diameter is approximately equal to the particle's Dmax) to another position that is 3.8 Å from a second arbitrary DR. Thus, the DRs are not directly connected to one another, but their placement is restricted to mimic a protein with 3.8 Å amino acid backbone spacing. DR movement continues throughout the simulated annealing protocol until Equation 3.15 is minimized. Some penalty designations also differ between DAMMIF and GASBOR.

The complete merged scattering curve, including both small and wide angle scattering data, for each apo polymerase was used for GASBOR shape modeling. Default values were used for all parameters, except the number of amino acid (aa) residues was set to match that of the protein (Taq – 832 aa, Klentaq – 554 aa, and Klenow – 605 aa).

Shape modeling with both DAMMIF and GASBOR was performed on the Louisiana Optical Network Initiative's (LONI) Eric2 high performance computing (HPC) Linux cluster, which is housed on Louisiana State University's Baton Rouge campus. For each X-ray scattering curve, 10 – 15 bead models were generated using serial processing allocated by a user generated Portable Batch System (PBS) script. Multiple bead models allow for evaluation of the stability of individual model features. For each model set, the models were compared, aligned, averaged, and then filtered to the average excluded volume of the models in the set using the

automated analysis algorithm implemented by DAMAVER [116, 134]. Any significantly deviating model (outlier) was excluded from the alignment and further analysis. During DAMAVER processing, the dummy atom model (DAM) packing radius was decreased to between 2.0 – 2.75 Å to increase the number of beads remaining in the final filtered model. Unless otherwise noted in this manuscript, the final shape model is the averaged and filtered model. For the apo polymerases, in addition to the automated processing of the individual bead models, each bead model was first superimposed with the corresponding crystal structure. Then, these superimposed models were averaged and filtered using the non-automated, interactive modes of DAMAVER and DAMFILT [116, 134].

Final shape models were superimposed with the corresponding crystal structure or ACM using the computer program SUPCOMB [116, 135], which is the same program used to align the individual bead models during the automated DAMAVER processing. SUPCOMB aligns arbitrary low and high resolution structures by minimizing a parameter termed the normalized spatial discrepancy (NSD). The NSD describes the sum of the distances between each bead in the first structure and all beads in the second structure and vice versa, normalized against the average distances between adjacent beads for the two structures. The mirror image of a structure exhibits the same X-ray scattering curve, thus each bead model may be an enantiomorph of the actual structure. During alignment of a shape model with a high resolution structure, SUPCOMB also examines enantiomorphs to minimize the NSD [116, 135].

3.5.2.3.4 Preparation of Atomic Coordinate Models (ACMs)

Atomic coordinates from PDB entries 4KTQ [39] and 1KLN [37] were used to prepare the atomic coordinate models (ACM) of the polymerase-DNA complexes for comparison with experimental X-ray scattering data. The various models created include both a polymerization

and an editing mode version of each of the 4 polymerase + DNA complexes noted in Table 3.3 (thus, a total of 8 different ACMs). Coordinates for the isolated DNA constructs, ds-63/63mer and pt-63/70mer, wound in perfect B-DNA helical conformations as shown in Figure 3.11 were generated from the respective DNA sequences using the "make-na" online server [136], which automates the Nucleic Acid Builder (NAB) computer code [137]. NAB is a high-level programming language for simulating both regular and irregular nucleic acid structures [137].

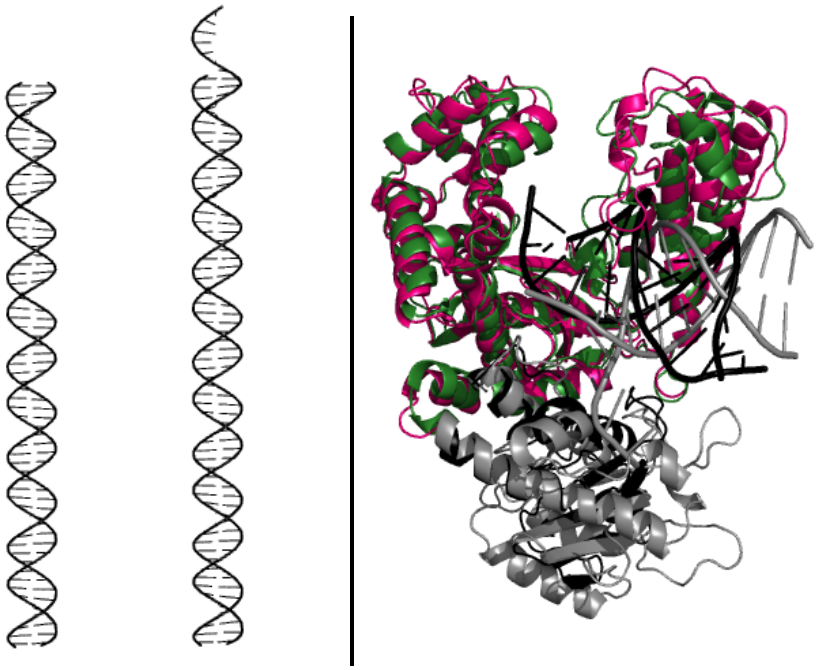


Figure 3.11. Protein and DNA constituents of the polymerase-DNA complex ACMs (atomic coordinate models) used for alignments with the 3D solution scattering shape models. *Left and middle:* ds-63/63mer and pt-63/70mer DNA molecules simulated via the "make-na" web server [136]. *Right:* Alignment of the polymerase domains of Klentaq (pink) and Klenow (green) from PDB entries 4KTQ and 1KLN, respectively (Cα RMS deviation equals 1.6 Å). The DNA components of these crystal structures and the proofreading domains of Klentaq and Klenow, colored black and gray, respectively, were not included in the alignment. In this figure, the DNA molecules are shown on a different size scale than the proteins.

4KTQ is a binary co-crystal structure of Klentaq complexed with a 12/16mer primed-template DNA substrate bound in the polymerization mode. The 3' terminal base of the primer strand is a chain terminating ddCTP that was incorporated by the polymerase. Coordinates corresponding to the three single-stranded nucleotides at the 5' terminus of the template strand are absent due to disorder and lack of interpretable electron density. The protein component of 4KTQ is also missing coordinates for residues 281-293 (full-length Taq numbering) at the N-terminus [39]. However, the Klentaq fragment overexpressed in our laboratory and used for these X-ray scattering experiments is comprised of residues 281 – 832 with a two amino acid N-terminal extension (met-gly) [91].

The 1KLN editing mode complex contains protein atomic coordinates for amino acids 324 – 582, 585 – 600, and 609-928 (missing residues all belong to the thumb subdomain), and DNA coordinates corresponding to a 10 nucleotide template strand and a 13 nucleotide primer strand. Three single-stranded nucleotides located at the 3' primer terminus are bound within the Klenow editing domain [37].

PyMOL [138] was used to visualize and manipulate all macromolecules during preparation of the polymerization and editing mode polymerase + DNA atomic coordinate models. As depicted in Figure 3.11, Klentaq and Klenow were first superimposed in PyMOL by aligning the Cα backbone of Klentaq's polymerase domain (residues 420 – 832) against the homologous region in Klenow (residues 516 – 928) [33]. To prepare the polymerization mode complexes, the ds-63/63mer and pt-63/70mer were manually superimposed with the 4KTQ DNA substrate. Then the original DNA substrate was replaced with each simulated DNA construct, and new coordinate files were written to yield Klentaq + DNA polymerization mode ACMs, dsKTQ-pol and ptKTQ-pol. Since the protein constituents were aligned, we then replaced the

original 1KLN DNA substrate coordinates with those of the DNA molecules superimposed with 4KTQ to generate the Klenow polymerization mode complexes, dsKLN-pol and ptKLN-pol. Following alignment of the ds and pt DNA macromolecules with the 1KLN DNA substrate, the editing mode complex models (ptKLN-edit, dsKLN-edit, ptKTQ-edit, and dsKTQ-edit) were prepared via the same procedure used to generate the polymerization mode models.

As indicated by the lack of interpretable electron density corresponding to the single-stranded nucleotides of the 5' template overhang in 4KTQ and 1KLN, these ss nucleotides are flexible and disordered [37, 39]. Depending on the binding mode, polymerization versus editing, 7 or 10 nucleotides, respectively, comprise the ss template overhang of the 63/70mer. Each overhang is long enough to extend outside of the polymerase envelope, where it may contribute to the X-ray scattering pattern. (The 63/63, when bound in the editing mode, also has a short ss 5' template overhang, but it is not long enough to escape the polymerase envelope.) Therefore, the positioning of the 5' template overhang in our ACMs is important for comparative purposes. To orient the template overhang we used cues from crystal structures of the closed ternary complex of Klentaq (PDB ID 3KTQ) [39] and two other family A polymerases, *Bacillus stearothermophilus (Bst)* and T7 DNA polymerases bound to DNA (PDB ID's 4BDP and 1T7P) [39, 42, 43]. These 3D structures, depicted in Figure 3.12, show 1 – 2 nucleotides of the template overhang flipped out of the duplex base stacking arrangement by approximately 90 - 180° and indicate that the ss 5' template overhang passes across the fingers subdomain, instead of through the cleft between the fingers and thumb as previously suggested [37]. Crosslinking studies performed on Klenow upon association with a DNA substrate support a pathway for the 5' ss template overhang across the fingers subdomain as suggested by crystallographic observations [49].

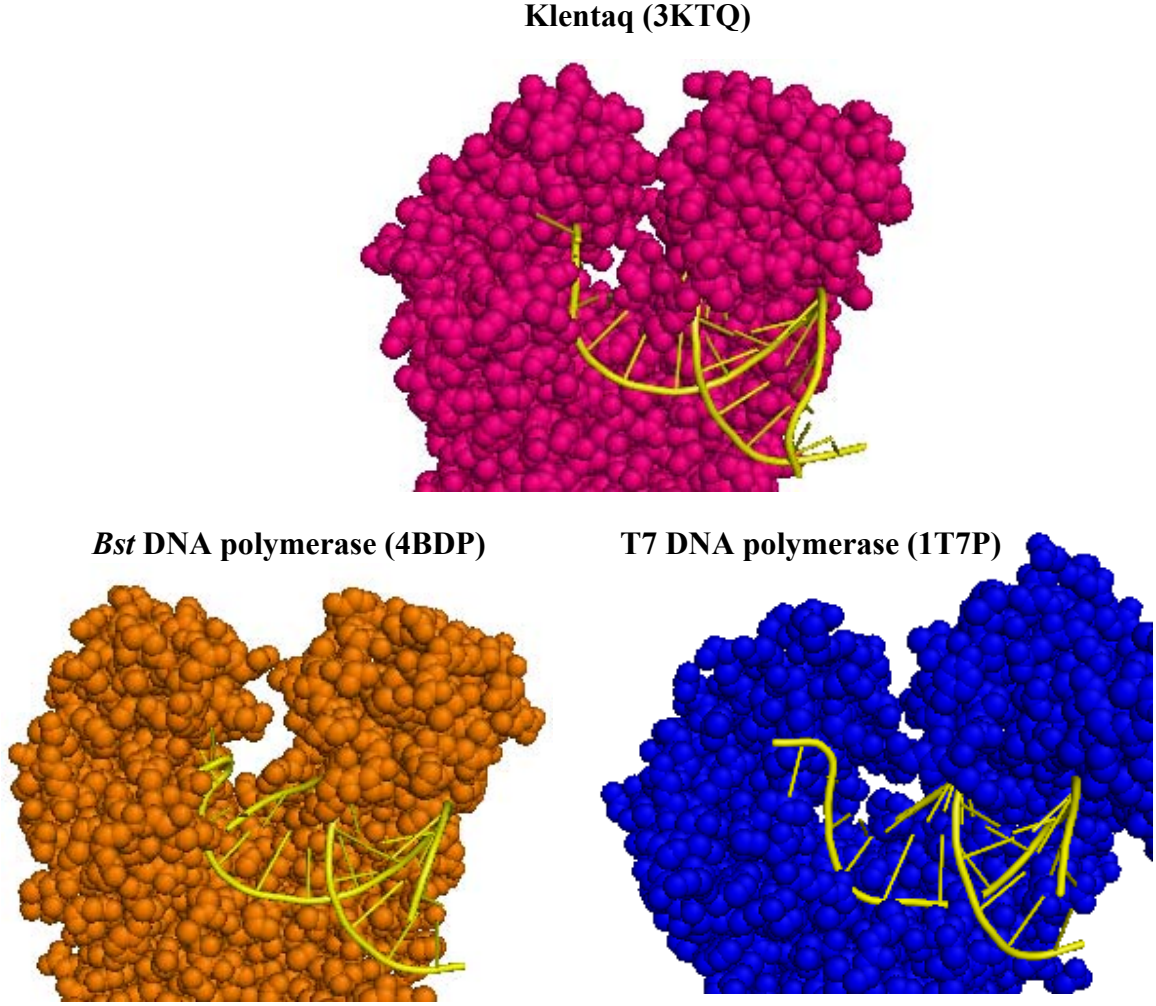


Figure 3.12. Atomic resolution structures of DNA polymerases belonging to the Pol A family suggest that the 5' ss template overhang enters the polymerase active site from the fingers subdomain, instead of passing through the polymerase cleft between the fingers and thumb subdomains. DNA substrates are depicted in a yellow cartoon representation, while the polymerase domains of Klentaq, *Bst* Pol I, and T7 DNA polymerase are shown in spacefill representation. In each of these structures, 1 – 2 ss nucleotides at the 5' end of the template strand are rotated out of base-stacking plane and oriented towards the fingers subdomain.

As shown in Figure 3.11, the single-stranded portion of the originally simulated 63/70mer maintains the B-form helical nature of the duplex. Necessary positional modifications of the ss template nucleotides, guided by other studies [39, 42, 43, 49], were thus made to reorient the overhang across the fingers subdomain, while minimizing steric conflicts with protein residues.

For the most part, major changes were localized to only the first few ss nucleotides downstream of the duplex, while the subsequent overhang nucleotides retained the helical, base stacked conformation of the simulated B-DNA construct.

3.5.2.3.5 CRYSOL Analysis with Crystal Structures and ACMs

The program CRYSOL [121] calculates the theoretical scattering curve from atomic coordinates, then alters the volume and hydration contrast parameters to obtain the best fit to the experimental scattering curve. CRYSOL (versions 2.3 and 2.6) calculations were performed as described by Svergun, et al. [121] for all crystal structures (1TAQ [30], 1CMW [104], 1KTQ [33], and 1KFD [34]) and ACMs examined. For SAXS curves ($q_{max} < 0.3$), program default values were used for all parameters except the scattering q maximum, which was set to coincide with the q maximum of the experimental data sets. To obtain CRYSOL simulated scattering curves with q_{max} greater than 0.3 $Å^{-1}$, version 2.6 was used and the maximum order of harmonics, order of the Fibonacci grid, and number of angular points were respectively set to 50, 18, and 201 or 256. The solvent electron density for simulated scattering curves compared with the SSRL-Jun07 experimental scattering data, which were collected in a glycerol buffer, was 0.343 $e^-/Å^3$.

Several crystal structures for Klenow polymerase exist in various functional states. Conformational variations among the different states for Klenow are not significant enough to be easily detectable by SAXS, and CRYSOL calculations with different Klenow structures yield equivalent results (data not shown). Rigid body rotations of the position of the 5' nuclease domain of Taq relative to the Klentaq domain were performed using the program Insight II (Accelrys, Inc.).

To obtain the Rg and Dmax values calculated from the crystal structures and ACMs, the simulated scattering curves were analyzed directly before they were fit to experimental data. In these cases, initialized parameters for 1) contrast of the solvation shell, 2) average atomic radius, and 3) excluded volume were used. The solvation shell contrast is initialized at 0.03 $e^-/Å^3$, and the average atomic radius and excluded volume are determined from the macromolecular composition.

CHAPTER 4

FULL-LENGTH TAQ DNA POLYMERASE ASSUMES AN ELONGATED CONFORMATION IN SOLUTION*

4.1 Introduction

Understanding the conformation of an enzyme in solution is an essential aspect of the comprehensive understanding of the overall structure and function of that enzyme. The spatial positioning of the 5' nuclease domain of Taq DNA polymerase has been the subject of debate, and crystal structures of the full-length protein have been reported with the 5' nuclease domain in two different positions relative to the Klentaq domain [30, 41, 104]. The two structures are shown in Figure 4.1. In one structure the 5' nuclease domain extends directly out into solution, while in the other this domain is folded up against the Klentaq domain.

The two reported conformations for Taq polymerase have potentially different functional consequences. In the elongated conformation of the polymerase, the polymerization and 5' nuclease active sites are separated by approximately 70 Å [30, 104]. In order for a single polymerase molecule to catalyze simultaneously nucleotide incorporation and 5' nuclease activities in close proximity on the same piece of DNA, the polymerase would need to adopt a more compact conformation which would bring the two active sites into relative proximity [104]. While the compact conformation observed in the Urs, et al. crystal structure is not ideally oriented for such simultaneous catalysis, the two active sites are in closer proximity than

---

* Reprinted from The Journal of Biological Chemistry, 278, A.M. Joubert, A.S. Byrd, and V.J. LiCata, Global conformations, hydrodynamics, and X-ray scattering properties of *Taq* and *Escherichia coli* DNA polymerases in solution, 25341-25347, Copyright 2003, with permission from The American Society for Biochemistry and Molecular Biology, Inc.

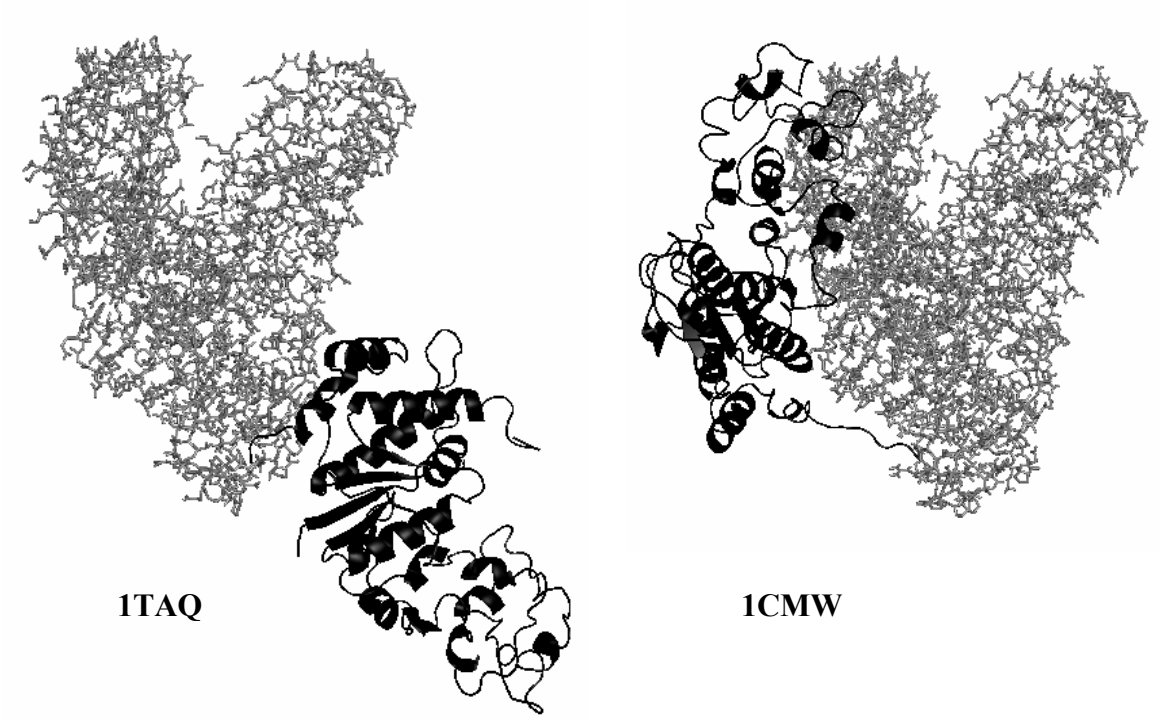


Figure 4.1. Alternate X-ray crystal structures of full-length *Taq* DNA polymerase [30, 104]. The Klentaq domains (light gray) of the two structures are equivalent, but the position of the 5' nuclease domain (black) differs significantly.

in the extended structure [104]. The larger separation between active sites in the elongated conformation of the polymerase does not preclude binding of the same DNA to the two active sites, however, it would involve more distant spacing along the DNA between the two active sites. It is also certainly possible that under certain conditions the polymerase can switch between elongated and compact conformations.

Analytical ultracentrifugation and small angle X-ray scattering provide different but complementary information on the size and shape of macromolecules in solution. The structural information provided by these techniques is quite "low resolution," but for large proteins, where NMR methods are only in their infancy, these techniques often provide some of the only solution structural information about a protein. In this study, the basic hydrodynamic

properties ($s_{20,w}$ and partial specific volume) and X-ray scattering properties (radius of gyration) of full-length Taq and *E. coli* Pol I DNA polymerases, and their Klentaq and Klenow subfragments have been characterized. We have also performed structure based calculations of these same parameters using software developed by Garcia de la Torre and associates (for hydrodynamic parameters) and by Svergun and associates (for X-ray scattering parameters) [103, 121]. In addition to presenting the basic hydrodynamic and X-ray scattering properties of these proteins, the data indicate that in the absence of DNA both full-length Taq and full-length Pol I are in an elongated conformation in solution, with their 5' nuclease domains sticking out into solution. Additionally, when bound to a matched primer-template piece of DNA or to ddATP, Taq polymerase remains in an elongated conformation.

4.2 Results and Discussion

4.2.1 Sedimentation Coefficients

Sedimentation coefficients were measured for all four polymerases: Taq, Klentaq, *E. coli* Pol I, and Klenow. Measurements were made at 20°C and were corrected to $s_{20,w}$ values using measured solvent densities and viscosities. Figure 4.2 shows representative sedimentation velocity data for Taq polymerase. Values of $s_{20,w}$ were also calculated from crystal structures using the program HYDROPRO from the laboratory of Garcia de la Torre [103]. This program uses shell modeling algorithms to predict hydrodynamic properties directly from a PDB file. The measured and calculated $s_{20,w}$ values for the polymerases are reported in Table 4.1.

The two different crystal structures of full-length Taq yield two different calculated $s_{20,w}$ values, as might be expected. The more compact form of the protein would sediment faster than the more elongated form, and its calculated $s_{20,w}$ is thus larger than that of the elongated form. Figure 4.3a graphically illustrates the relationships between the measured and calculated data.

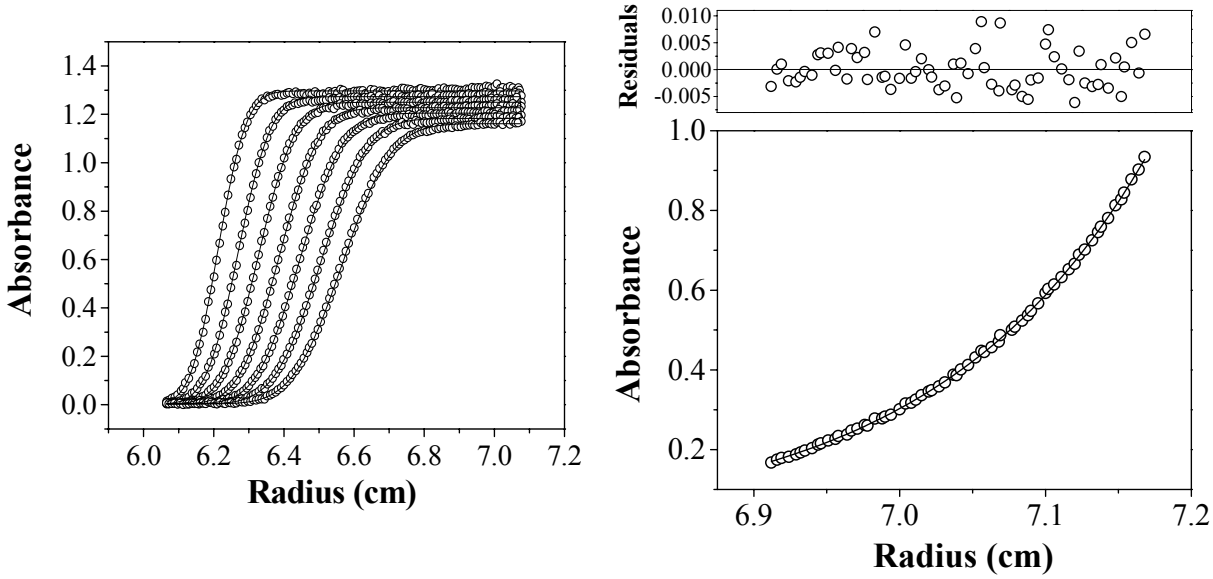


Figure 4.2. Representative velocity sedimentation data (left panel) and equilibrium sedimentation data (right panel) for full-length Taq DNA polymerase. The velocity sedimentation data shows the experimental data along with the fits to the boundaries using the program Svedberg [100, 101]. For visual clarity only a subset of the full set of boundaries is shown, and only half the data points in any single boundary are shown. The right plot shows the equilibrium sedimentation profile for Taq along with a single species fit. The residuals plot shown is for the single species fit to the equilibrium sedimentation data. Similar experiments were performed for Klentaq, *E. coli* Pol I, and Klenow polymerases, as described in the text. Results are reported in Table 4.1 and in the text.

Table 4.1. Hydrodynamic properties of the polymerases.

Protein	**$s_{20,w}$ (experimental)**	**$s_{20,w}$ (HYDROPRO)**	**R_s[c] (experimental)**	**R_s (HYDROPRO)**	**$D_{20,w}$ (experimental)**	**$D_{20,w}$ (HYDROPRO)**
	x10^{-13} s	**x10^{-13} s**	**Å**	**Å**	**x10^{-7} cm^2/s**	**x10^{-7} cm^2/s**
Taq	5.05 ± 0.04[a]	4.97$^{(1TAQ)}$ 5.31$^{(1CMW)}$	43.0 ± 0.4	42.7$^{(1TAQ)}$ 39.9$^{(1CMW)}$	5.10	5.01$^{(1TAQ)}$ 5.35$^{(1CMW)}$
Klentaq	4.08 ± 0.03[a]	4.01	35.3 ± 0.3	35.1	6.06	6.09
Klenow	4.59 ± 0.06[b]	4.54	35.3 ± 0.4	35.7	6.06	5.99
Pol I	5.29 ± 0.03[b]	N/A	44.4 ± 0.3	N/A	4.81	N/A

[a] Average of four measurements.
[b] Average of three measurements.
[c] Calculated from $s_{20,w}$ as described under "Results and Discussion."

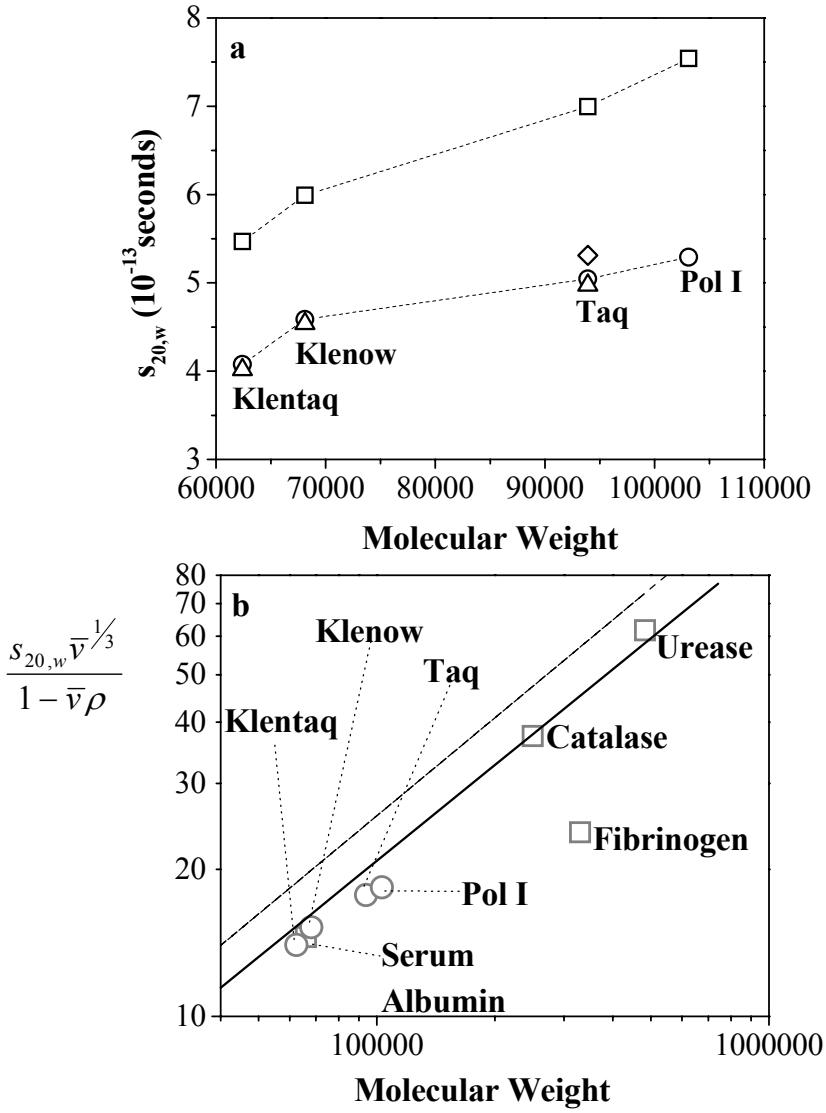


Figure 4.3. Plots of sedimentation coefficients ($s_{20,w}$ values) versus molecular weight for each of the four polymerases. The upper panel shows $s_{20,w}$ values plotted directly against protein molecular weights. Open circles show the experimental values for each of the polymerases. Molecular weights of the polymerases are 62.4 kDa for Klentaq, 68.1 kDa for Klenow, 93.9 kDa for Taq, and 103.1 kDa for Pol I. The line connecting the experimental data points is only for visual clarity. The open triangles show the $s_{20,w}$ values calculated from the crystal structures using the program HYDROPRO [103] for Klentaq (1KTQ [33]), Klenow (1KFD [34]), and elongated Taq (1TAQ [30]). The open diamond shows the calculated $s_{20,w}$ for the compact Taq crystal structure (1CMW [104]). Also shown, for comparative purposes, are the calculated $s_{20,w}$ values for spherical particles of the same molecular weights and partial specific volumes as the polymerases (open squares). The lower panel shows an alternate perspective on the hydrodynamic data. Here the experimental values of $s_{20,w}$ for each of the polymerases are shown along with representative data for several other globular and not so globular (e.g. fibrinogen) proteins. The upper solid line shows the $(s_{20,w}\bar{\nu}^{1/3})/(1-\bar{\nu}\rho)$ versus molecular weight dependence for perfect spheres, while the lower line shows this dependence for a set of globular proteins.

The measured and calculated $s_{20,w}$ values agree well for Klentaq and Klenow polymerases, and for the elongated form of full-length Taq. The measured $s_{20,w}$ value for full-length Taq deviates from the $s_{20,w}$ calculated for the elongated crystal structure by 1.6% and deviates from the $s_{20,w}$ calculated for the compact crystal structure by 5.1%. This is a small difference, but it is well outside experimental error, and its significance can be illustrated by noting that the classic T to R conformational transition for aspartate transcarbamylase (ATCase) is associated with a 3.6% change in its measured $s_{20,w}$ value [139]. The deviations between the measured and calculated $s_{20,w}$ values for Klentaq and Klenow are 1.7% and 1.1% respectively. It is also notable that the measured $s_{20,w}$ value for full-length *E. coli* Pol I is also smaller than the calculated $s_{20,w}$ for the compact conformation of Taq. Since *E. coli* Pol I is ~9 kDa larger than Taq, this further supports the interpretation that Taq is in the elongated conformation. Further, it argues that *E. coli* Pol I is also not in a compact conformation.

Figure 4.3a also shows the predicted $s_{20,w}$ values for spherical particles of the same molecular weight and partial specific volume ($\bar{\nu}$) for each of the polymerases. These values are not precisely linear due to the slight differences in $\bar{\nu}$ among each of the polymerases (see below), but do illustrate, for comparative purposes, the increase in $s_{20,w}$ expected for particles that are increasing in mass but not changing shape. Comparison of the measured $s_{20,w}$ values with these spherical values shows that the full-length polymerases deviate from the spherical limit more so than the Klenow and Klentaq polymerases. Figure 4.3b illustrates this same point in a different way. It can be shown by rearrangement of the Svedberg equation that the $s_{20,w}$ values for anhydrous spherical particles are proportional to the 2/3 power of their molecular weight [98]. The $s_{20,w}$ values for real globular proteins follow a similar proportionality relationship versus molecular weight [98]. Here both full-length polymerases are seen to

deviate further from the mean dependence for globular proteins than do their large fragment counterparts, indicating that they are more elongated than their large fragment counterparts.

Effective Stokes radii (Rs) for the polymerases can be calculated from the measured $s_{20,w}$ values using the equation: $Rs = MW(1- \bar{\nu}\rho)/6N\pi\eta s_{20,w}$ where MW is the molecular weight, $\bar{\nu}$ is the partial specific volume, ρ is the solvent density, N is Avogadro's number, and η is the solvent viscosity [140]. Such calculated Stokes radii report the effective spherical radius of a particle having a particular measured $s_{20,w}$ value. They are instructive in reflecting, in angstroms, the relative hydrodynamic sizes of the proteins. The diffusion coefficient ($D_{20,w}$) for each polymerase can also be calculated from the experimental $s_{20,w}$ and the molecular weight using the Svedberg equation: $D_{20,w} = s_{20,w}RT/MW(1-\bar{\nu}\rho_w)$, where R is the gas constant, T is the Temperature in Kelvin, MW is the molecular weight of the polymerase predicted from the amino acid sequence, $\bar{\nu}$ is the partial specific volume of the polymerase, and ρ_w is the density of water. These values are reported in Table 4.1 along with Rs and $D_{20,w}$ values calculated from the crystal structures using HYDROPRO.

4.2.2 Equilibrium Sedimentation and Partial Specific Volumes

In order to insure that the polymerases were monomeric and of high purity, and in order to measure directly the partial specific volumes of each of the polymerases, equilibrium sedimentation experiments were performed. Equilibrium sedimentation runs for all the polymerases fit well to a single species (representative data for Taq polymerase is shown in Figure 4.2). Molecular weights for the polymerases measured by equilibrium sedimentation deviated from their known molecular weights by 1% for Klentaq, 2.6% for Taq, 6.3% for Klenow, and 7.5% for *E. coli* Pol I. In all cases the equilibrium sedimentation determined

molecular weights were slightly lower than known molecular weights, further indicating the absence of higher order oligomers.

Equilibrium sedimentation was also used to measure directly the partial specific volumes ($\bar{\nu}$) for each of the polymerases. Values of $\bar{\nu}$ for the four proteins were both calculated from the primary structure, and were measured directly using differential equilibrium sedimentation in H_2O versus D_2O [102]. Figure 4.4 shows representative data from the H_2O/D_2O experiments, and $\bar{\nu}$ values are shown in Table 4.2. Calculated $\bar{\nu}$'s have been found to be quite adequate for use in most ultracentrifugation studies, and it is becoming increasingly rare to measure them directly. Small changes in the $\bar{\nu}$ do not propagate into significant effects on the values of the measured $s_{20,w}$'s. Small changes in the $\bar{\nu}$ do, however, have quite large effects on the calculated $s_{20,w}$ values estimated using the program HYDROPRO, and this is the primary reason we empirically verified the computational $\bar{\nu}$'s. It can be seen in Table 4.2 that the calculated and measured values are quite similar. The empirical determinations of the $\bar{\nu}$'s verify that there are no unusual properties of these thermophilic proteins that might skew the calculation of their $\bar{\nu}$'s relative to the non-thermophilic polymerases.

The H_2O/D_2O determinations themselves are not error free, of course, and contain small error contributions propagated from the solution density determinations, from completeness of the H_2O-D_2O exchange, et cetera. Therefore, and because the calculated and measured values are already similar, we have used the average of the calculated and measured values (Table 4.2) in all calculations requiring a value for $\bar{\nu}$.

It is notable that the similarity between the $\bar{\nu}$ values for the mesophilic and thermophilic polymerases suggest that there are no significant differences in the relative packing densities of the different polymerases. This finding is consistent with a recent computational study that

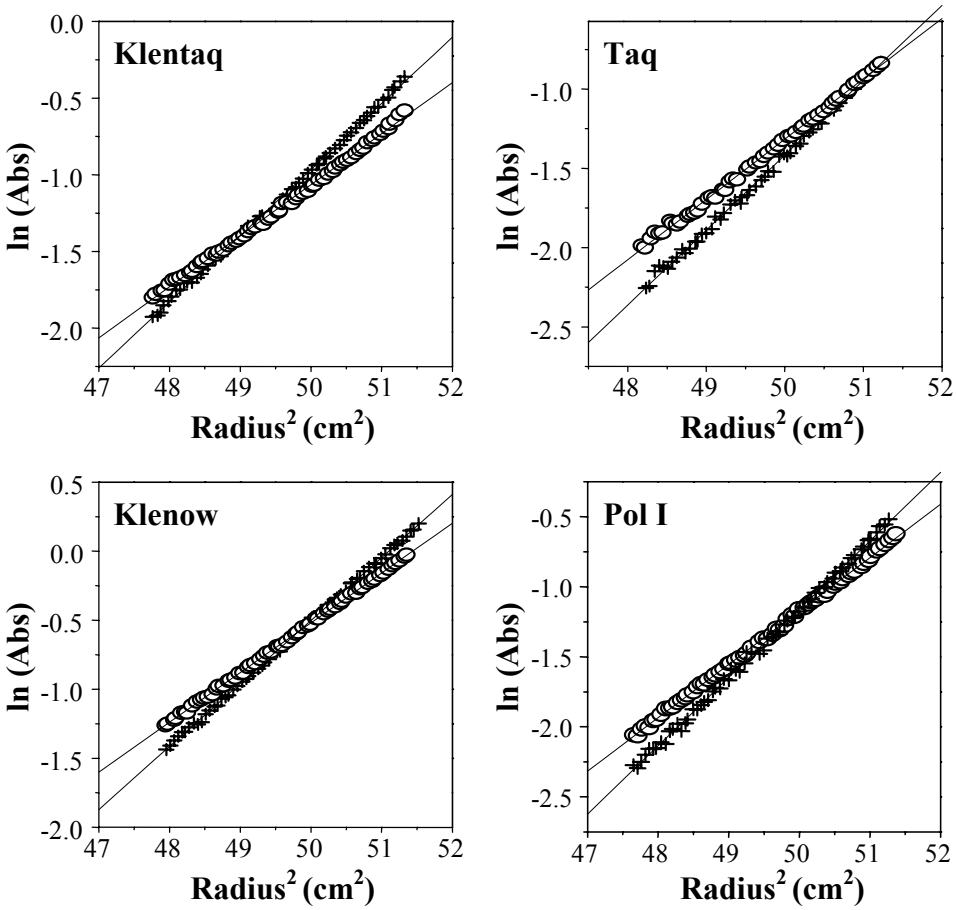


Figure 4.4. D_2O/H_2O sedimentation equilibrium experiments conducted for each of the polymerases. For each plot the open ovals are equilibrium sedimentation runs in D_2O and crosses (+) are equilibrium sedimentation runs in H_2O. Details are described in Chapter 3. Partial specific volumes are determined from the ratio of the D_2O/H_2O slopes using the method of Edelstein and Schachman [102]. Results are reported in Table 4.2.

Table 4.2. Partial specific volumes ($\bar{\nu}$) of the polymerases.

Polymerase	**Predicted $\bar{\nu}$** [a]	**Measured $\bar{\nu}$**	**Mean $\bar{\nu}$**
		ml/g	
Klentaq	0.744	0.732	0.738
Klenow	0.743	0.717	0.730
Taq	0.744	0.743 ± 0.043[b]	0.744
Pol I	0.742	0.740	0.741

[a] Calculated using the SEDNTERP software program.
[b] Average ± standard deviation from five experiments. The $\bar{\nu}$ values for the other polymerases were obtained from single experiments.

compared $\bar{\nu}$ values calculated from the crystal structures of a large set of mesophilic and thermophilic proteins and found no differences in relative packing densities [141].

4.2.3 Small Angle X-ray Scattering (SAXS)

SAXS experiments on each of the polymerases were performed using synchrotron radiation at several different protein concentrations under the same solution conditions as used to perform the hydrodynamics experiments described above. Guinier plots for each of the polymerases are shown in Figure 4.5. Fits of the full scattering curves using the program GNOM [119] are shown in Figure 4.6. The measured values for the radius of gyration (Rg) for each polymerase are listed in Table 4.3 along with the Rg values calculated from the different crystal structures. Both the programs HYDROPRO [103] and CRYSOL [121] were used to calculated Rg values from the crystal structure data. Unlike the data for the other polymerases, Guinier plots for *E. coli* Pol I consistently exhibited significant curvature throughout their entire q range at a variety of different protein concentrations and at all SAXS experimental stations used during this study. An example of this curvature is shown in the Guinier plot in Figure 4.5. Because of the curvature, *E. coli* Pol I scattering data were not analyzed for Rg values either by Guinier analysis or GNOM analysis.

The SAXS experiments shown also provide evidence that Taq is in an elongated conformation. As seen from the data in Table 4.3, the measured Rg value for full-length Taq agrees most closely with the Rg value calculated from the elongated structure. It should be noted, however, that the Rg data for full-length Taq, like the hydrodynamic data, actually suggests that the solution conformation is in between the two crystal structure conformations. This can be seen by noting that the measured Rg values for Klentaq and Klenow are both larger than the Rg values calculated from their crystal structures. This is a commonly observed

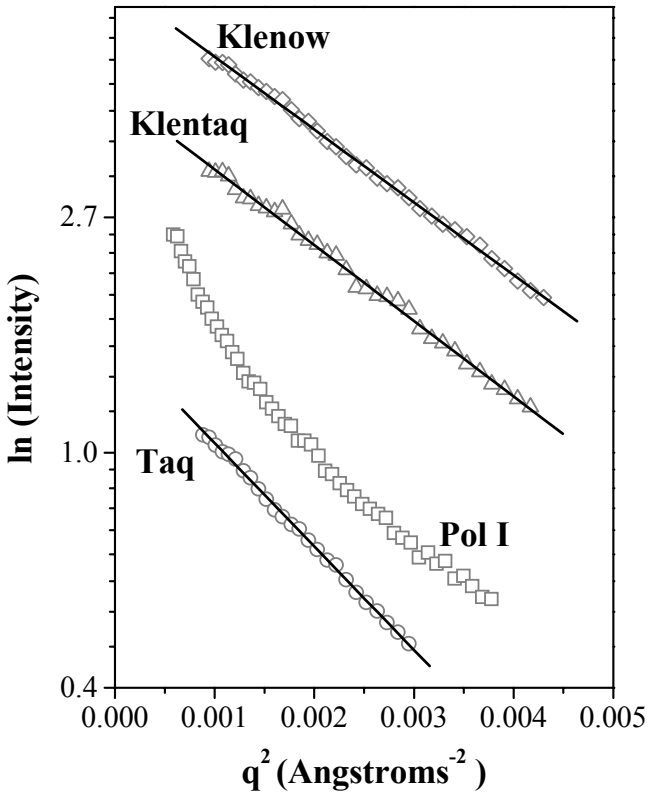


Figure 4.5. Guinier plots [117] of the natural log of the small angle X-ray scattering intensity versus q^2 for each of the DNA polymerases. Open circles show the data for full-length Taq, open triangles show data for Klentaq, diamonds show data for Klenow, and open squares show data for *E. coli* Pol I. Pol I shows considerable curvature in the Guinier plot and so was not analyzed for an Rg value. The fitted lines used to determine the Rg's for Taq, Klentaq, and Klenow reported in Table 4.2 are shown. Protein concentrations are 5.1 mg/ml for Taq, 3.8 mg/ml for Klentaq, 6.3 mg/ml for *E. coli* Pol I, and 5.3 mg/ml for Klenow. The data for Klentaq and Klenow are in the original ln (intensity) units. The data for Taq and Pol I have been displaced on the y-axis by multiplying the original intensity values by factors of 0.25 and 10 respectively, for visual clarity.

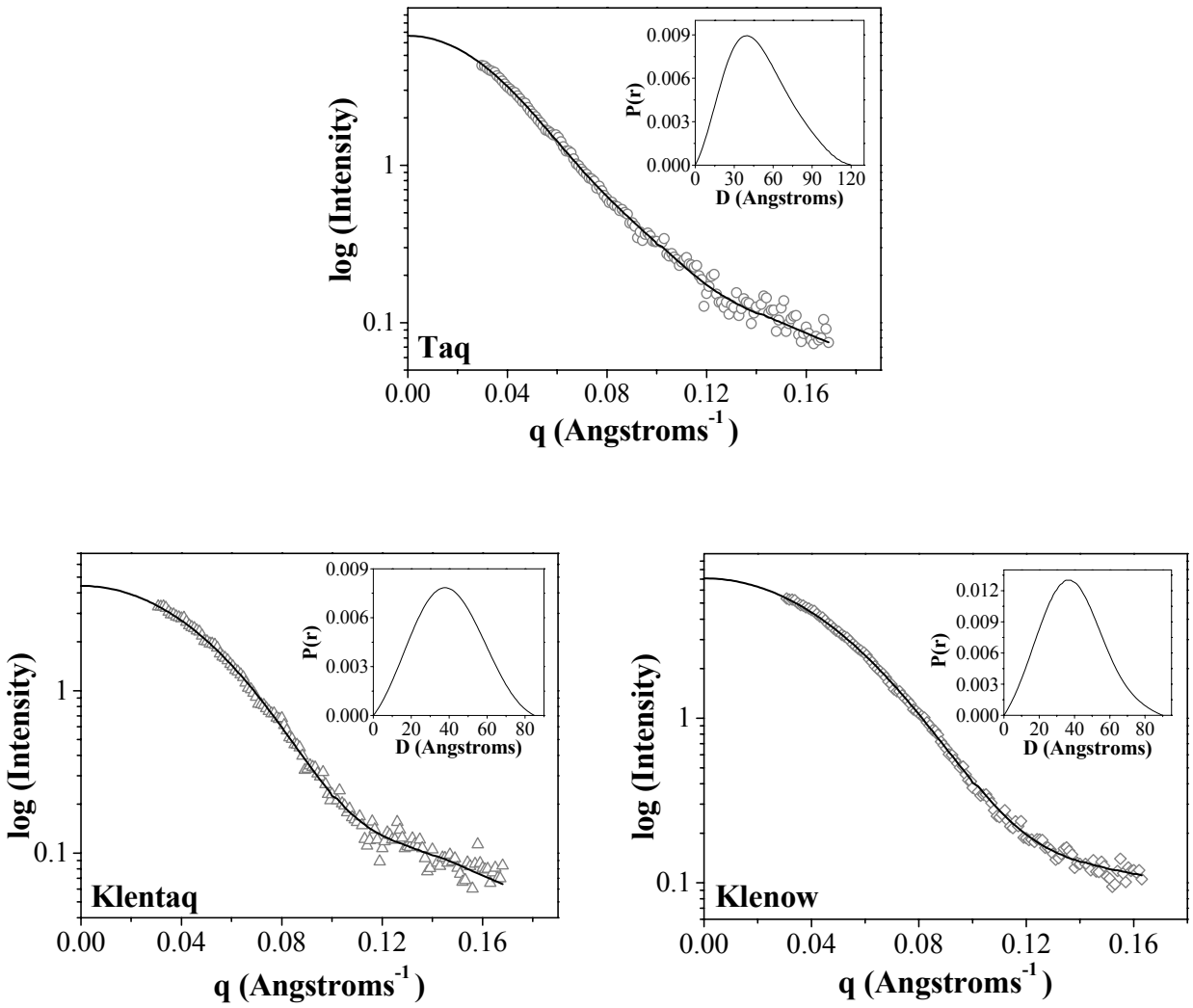


Figure 4.6. GNOM [119] fits of the log intensity versus q scattering profiles of Taq, Klentaq, and Klenow polymerases. Inset graphs show the P(r) distance distribution function versus Dmax for the best fit to each polymerase. Protein concentrations are 5.1 mg/ml for Taq, 3.8 mg/ml for Klentaq, and 5.3 mg/ml for Klenow.

Table 4.3. Measured and calculated radii of gyration for the polymerases.

Protein	Rg (Guinier)[a]	Rg (GNOM)[b]	Rg (experimental mean)	Rg (HYDROPRO)	Rg (CRYSOL)
			Å		
Taq	38.7 ± 1.0	37.9 ± 0.2	38.3 ± 1.2	37.5(1TAQ) 31.2(1CMW)	37.7(1TAQ) 31.1(1CMW)
Klentaq	31.2 ± 0.9	30.2 ± 0.1	30.7 ± 1.0	27.1	27.3
Klenow	31.2 ± 0.5	29.8 ± 0.1	30.5 ± 0.6	28.0	27.7

[a] Mean of five measurements for each protein.
[b] Value from one data set using data to high q.

discrepancy which is generally attributed to hydration effects [142]. This means that if the values could be precisely adjusted for hydration effects, the measured Rg for Taq would likely be in between the Rg values calculated for the two crystal structures. However, the measured Rg would still be closer to that predicted for the elongated structure. The calculated values do include standard hydration levels of 0.3-0.4 g of H_2O/g of protein (and similar hydration levels (~0.4 g/g) are predicted by calculations based on amino acid composition using the program SEDNTERP). However, adjustment of the calculated Rg values to account precisely for hydration is still an active area of research, and as such no unequivocal guidelines yet exist.

Figure 4.7 shows fits of the experimental scattering curves for full length Taq, Klentaq, and Klenow polymerases overlaid with predicted scattering curves computationally generated from their crystal structures. These fits, generated with the program CRYSOL, begin with a curve generated directly from the known crystal structures and then vary two parameters of the structure based simulation, the volume and the hydration contrast, to obtain the best correlation between the experimental curve and that predicted from the crystal structure [121]. The goodness of the fits for Klentaq, Klenow, and the elongated conformation of Taq are quite

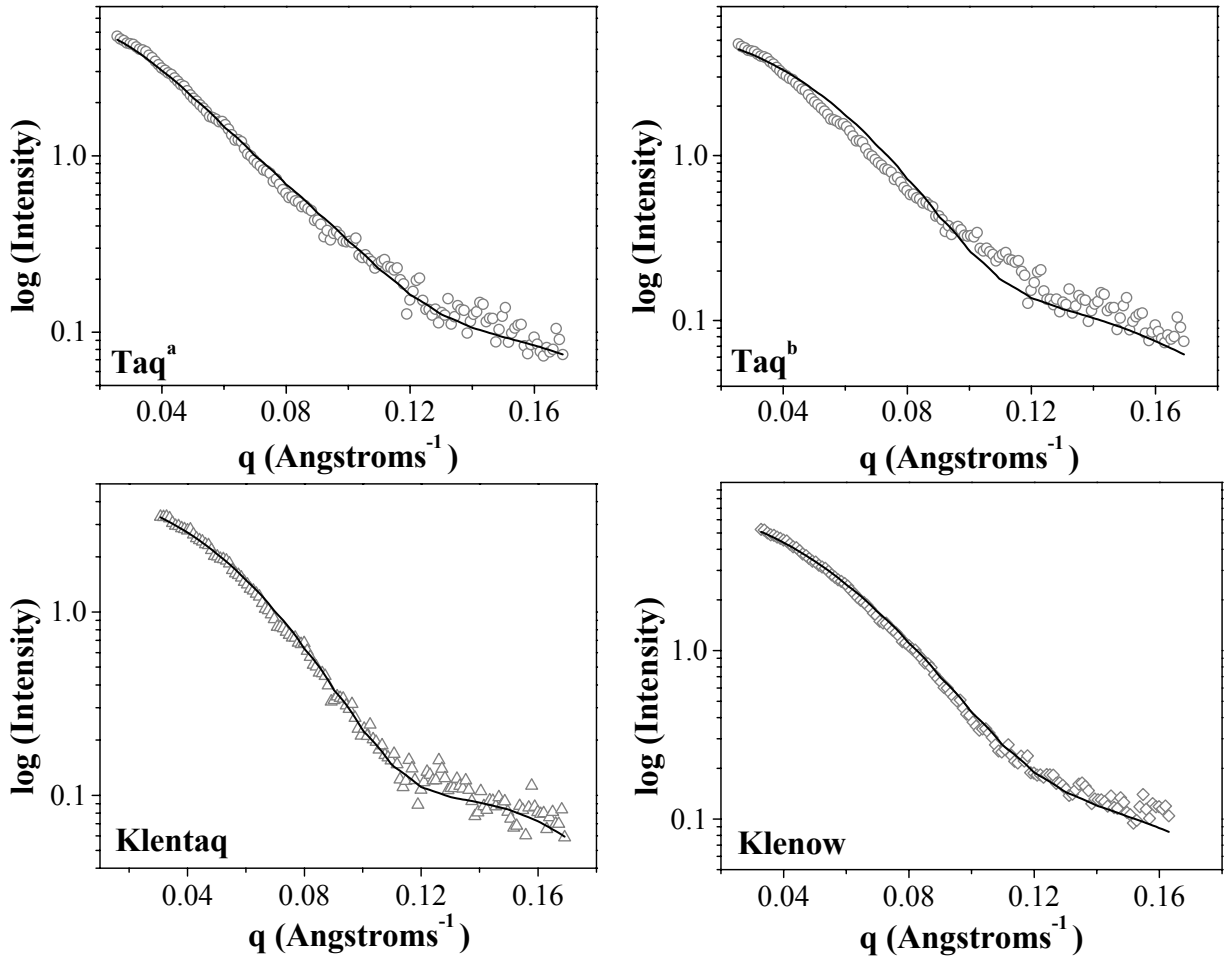


Figure 4.7. Correlations between crystal structure data and small angle scattering data. Fits of the experimental scattering data with simulated scattering data calculated from the two different crystal structures for full-length Taq polymerase (Taq^a = 1TAQ [30], Taq^b = 1CMW [104]), Klentaq polymerase (1KTQ [33]), and Klenow polymerase (1KFD [34]), performed using the program CRYSOL [121]. Open symbols show the experimental data, solid lines show the fitted simulated scattering data calculated from the crystal structures. Protein concentrations are 5.1 mg/ml for Taq, 3.8 mg/ml for Klentaq, and 5.3 mg/ml for Klenow. Chi (χ) values for the goodness of fit were 3.78, 6.05, 3.74, and 3.23 for the fits to Taq^a, Taq^b, Klentaq, and Klenow, respectively.

similar. Both the statistics (χ) and visual inspection of the fits in Figure 4.7 provide some of the most compelling evidence that Taq is best described as being in an elongated conformation.

The effects of simplistic, rigid body rotations of the position of the 5' nuclease domain of Taq were also examined. Inspection of the two different crystal structure positions of the 5' nuclease domain shown in Figure 4.1 shows them to be oriented approximately 180^{o} relative to each other. Movement of the 5' nuclease domain up to 45^{o} away from its elongated orientation in the 1TAQ structure results in relatively small changes in the goodness of fit to the experimental data using CRYSOL. Both slight improvements and slight decreases in the goodness of fit were observed in this range of movement. However, movement of the 5' nuclease domain more than 90^{o} away from its elongated orientation in the 1TAQ structure in any direction consistently resulted in significant decreases in the goodness of fit between the experimental and structure-based scattering profiles using CRYSOL. Possible alternate elongated Taq structures are not shown because the precision of the SAXS data at high q values do not allow for further distinction among the possible alternate elongated conformations, even if a more extensive three dimensional positional grid search were performed. This preliminary modeling exercise does, however, further reinforce the conclusion that the solution structure of Taq is more like the elongated crystal conformation than the compact conformation.

SAXS measurements of the Rg for full-length Taq polymerase were also carried out in the presence of ddATP and in the presence of a matched template-primer known to bind stoichiometrically to Taq under the conditions examined [81]. The potential effect of ddATP was examined because at least one ddNTP was present at high concentration in the crystallization conditions for 1CMW, the compact conformation of Taq [104]. Neither the dideoxynucleotide nor the matched DNA altered the measured Rg. Direct DNA binding studies

of Taq to this same DNA also indicate that the 5'nuclease domain is not involved in binding to matched DNA [81]. Recent studies from Dahlberg and associates have shown that 5' nuclease enzymatic activity of Taq is exhibited on specifically structured bifurcated gapped DNA duplexes [143, 144]. It is not yet known if both the polymerase and 5' nuclease domains of a single polymerase molecule can simultaneously act on or bind to one of these more complex duplex structures.

4.3 Concluding Discussion

In this study we have determined some of the basic hydrodynamic and small angle X-ray scattering properties of full length Taq and *E. coli* Pol I DNA polymerases and their Klentaq and Klenow large fragments. The measured $s_{20,w}$ values, partial specific volumes, Rs values, $D_{20,w}$, and Rg values have all been determined using established biophysical methods which yield relatively straightforward answers, adding new particulars to the body of data on these important enzymes. What is more equivocal, however, is answering the question: what is the conformation of Taq polymerase in solution? For example, as noted by Svergun and associates [142] direct comparisons between high resolution crystal structures and the low resolution information provided by SAXS must always be viewed with caution, and that even seemingly perfect agreement between the two methods would not be unequivocal proof of the equivalence of the crystal structure and the solution structure. Furthermore, the interrelated fields of calculating hydrodynamic [103] and X-ray scattering properties [121] from crystal structure data are both relatively new, and both still contain clear computational gaps which must be bridged between calculation and experiment (for example: precisely accounting for protein hydration) [103, 121]. With these and other caveats in mind, in this study we have found that every measured hydrodynamic and X-ray scattering property for Taq polymerase consistently

correlates with a more elongated conformation of the molecule. Further, when SAXS curves are simulated directly from the crystal structures and fit to the experimental data, the agreement between the experimental data and the elongated structure is clearly better. However, the exact/detailed elongated conformation of Taq in solution is certainly different from the exact elongated conformation in the crystal structure. The data here certainly indicate that Taq in solution is much more similar to the elongated crystal structure than the compact crystal structure, but the hydrodynamic and scattering data presented here do not provide enough structural detail to determine the detailed differences and similarities between the elongated crystal and solution structures. A central goal of this study was to characterize the conformation of isolated/apo Taq polymerase in solution (i.e. without bound DNA) in comparison to the two contrasting crystal structures available for isolated/apo Taq polymerase. It is certainly very possible that when bound to certain DNA, i.e. during nick-translation, that the polymerase could move to adopt a compact conformation with both the polymerase and 5' nuclease domains bound to the same DNA. Further studies are required to begin to explore such possibilities.

CHAPTER 5

X-RAY SCATTERING BASED SHAPE MODELING OF *TAQ* AND *E. COLI* DNA POLYMERASES IN SOLUTION: GLOBAL CONFORMATIONS AND DIFFERENTIAL FLEXIBILITY

5.1 Introduction

The hydrodynamic and scattering data presented in Chapter 4 clearly indicate that Taq DNA polymerase assumes an elongated conformation in solution, and that the measured global solution structure is very similar to that of the 1TAQ crystal structure. Although the scattering patterns simulated from the 3-dimensional atomic structures agree very well with the experimental SAXS curves for Klentaq and Klenow, the experimental Rg values suggest that Klentaq and Klenow are 2-3 Å larger than expected according to their respective crystal structures.

Small angle X-ray scattering curves ($q < 0.3$ Å^{-1}) like those presented in Chapter 4, however, typically contain only structural information concerning the overall shape or global conformation of the macromolecule (resolution = $2\pi/q < \sim 20$ Å). Svergun and colleagues suggest that scattering curves comprised of data within the wide angle scattering regime ($q >$ 0.3 Å^{-1} up to ~1.3 Å^{-1}) also encompass internal structural information pertaining to the fold of the scattering particle (resolution up to ~5 Å) [122]. The *ab initio* bead modeling program, GASBOR [122] is specifically suited to account for the entire scattering curve up to a scattering vector of ~1.3 Å^{-1} and can reconstruct higher resolution shape models than predecessor programs, such as DAMMIN [107]. Thus, to gain further insight into the conformation of the large fragments of *Taq* and *E. coli* Type I DNA polymerases and full-length Taq, we collected additional X-ray scattering data across both the small and wide angle scattering regimes (q range 0.01 – 0.77 for large fragments and 0.02 - 1.1 Å^{-1} for Taq) and created shape models of

the polymerases using GASBOR [122]. These higher resolution X-ray scattering models were compared to existing crystal structures, and we show that the solution scattering based models overlap the known crystal structures well (in particular the elongated crystal structure for full-length Taq), and reveal a distribution of flexible equilibrium conformations of the polymerases, which can explain the apparent 2-3 Å difference between the SAXS solution structures and the crystal structures.

5.2 Results and Discussion

5.2.1 Preliminary Analysis and Preparation of X-ray Scattering Curves for Shape Modeling

The scattering intensity for a particle in solution decays rapidly with increasing scattering angle, and within the wide angle regime the particle scattering intensity is just barely above that of the background. Thus, to collect high quality, low noise wide angle X-ray scattering (WAXS) data, we measured the scattering curves using a very stable high flux X-ray beam. All of the X-ray scattering data presented in Chapter 4 were collected on beamlines with relatively low flux and exhibited no statistically significant dependence on protein concentration or signs of X-ray induced radiation damage. For data collected with high flux beams, some macromolecules show such problems while others do not. For example, SAXS curves for Klentaq (see Figure 5.1) and WAXS data for full-length Taq, Klentaq, and Klenow (data not shown) measured via the high flux X-ray beam were unaffected by either phenomenon. As demonstrated in Figure 5.1, in response to high flux X-rays apo Klentaq exhibits a relatively concentration independent SAXS pattern throughout the entire range of assayed protein concentrations (0.77 – 6.15 mg/ml). Extrapolation of the GNOM-derived I(0)/c and Rg data in Figure 5.2 to infinite dilution yields molecular weight (from lysozyme calibration of I(0)/c as

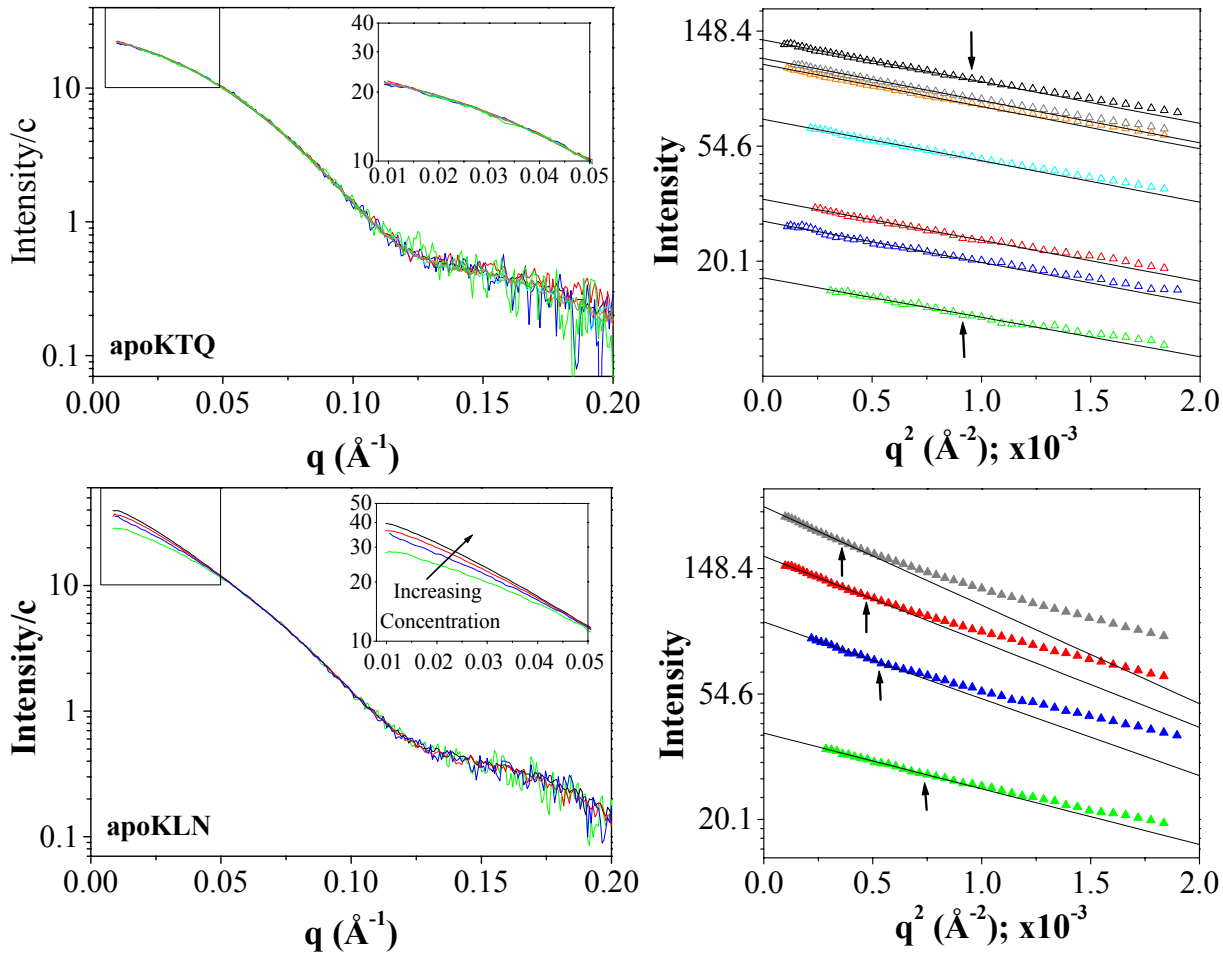


Figure 5.1. SAXS curves from apoKTQ *(top)* and apoKLN *(bottom)* concentration series are respectively representative of concentration independent X-ray scattering (apoKTQ) and aggregation (apoKLN). *Left:* Concentration normalized scattering intensity is plotted on a $\log_{10}$ scale as a function of momentum transfer (q). The range of q values affected by concentration dependent phenomena is restricted to very small scattering angles (boxed region), and is particle specific. The boxed regions are magnified in the inset graphs. *Right:* Corresponding Guinier plots of apoKTQ and apoKLN concentration series. The ordinate axis contains un-normalized scattering intensities plotted on a natural log scale. For each Guinier plot, concentration decreases from top to bottom. Concentrations vary from 0.77 – 6.15 mg/ml and are represented as different colors that correspond to the colors in the graphs on the left. Solid lines are the linear fits to the scattering data up to the arrows (marking the point where q_{max}*Rg approximately equals 1.3) and extrapolated to the abscissa.

described in Chapter 3) and Rg values for apo Klentaq of 56.5 kDa and 30.7 Å. Thus, Klentaq's Rg is not altered by the high flux X-rays (see Table 4.3), and its experimental molecular weight is within 10% of the molecular weight calculated from the known primary sequence (which is as good as can be expected for molecular weight determination via SAXS).

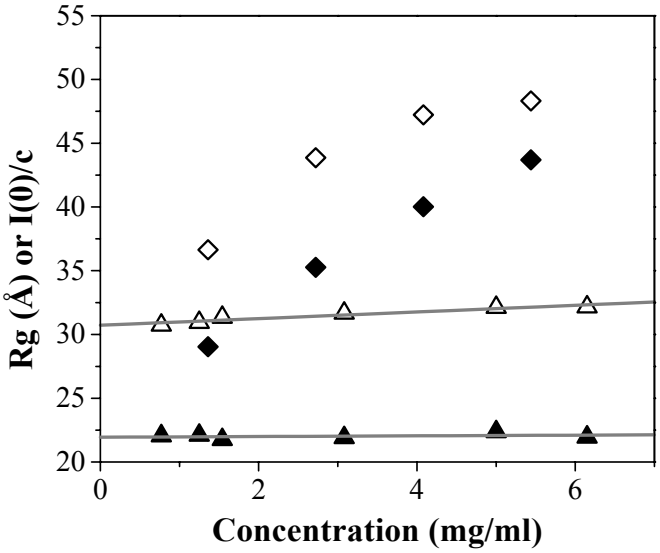


Figure 5.2. Concentration dependencies of GNOM derived Rg (open symbols) and I(0)/c (closed symbols) for Klentaq (triangles) and Klenow (diamonds). Errors are smaller than the symbols. Guinier analyses of the scattering curves yielded very similar Rg and I(0)/c values (data not shown). Solid lines are linear fits to the Klentaq data sets. Rg and I(0)/c values for Klenow exhibit nonlinear concentration dependence and were not fitted or extrapolated to infinite dilution.

Conversely, the small-angle scattering curves from full-length Taq and Klenow were adversely affected by the high flux X-rays. The Taq SAXS curves, which were measured under static conditions (i.e. the sample was not oscillated through the beam during X-ray exposure), exhibited a time-dependent scattering pattern indicative of X-ray induced aggregation and similar to the example depicted in Figure 3.7. On the other hand, aggregation of Klenow in response to high flux X-ray exposure (even though the sample was continuously oscillated

through the beam) was manifested as a concentration dependent pattern at very low scattering angles (see Figure 5.1). Within the low q regime affected by aggregation as shown in Figure 5.1, I(q) increases with increasing protein concentration, thus steepening the slope of the Guinier plot and enlarging the Rg and I(0)/c values in a nonlinear fashion (see Figure 5.2). At a scattering vector above approximately 0.05 $Å^{-1}$, all of the Klenow scattering curves superimpose, and the scattering intensity at subsequent q values is independent of concentration, which indicates that the internal structure of Klenow is not affected by the aggregation. The observed aggregation of Klenow and Taq seems to be have been caused by the high flux X-ray exposure, as no evidence of aggregation of either polymerase was obtained from 1) low flux SAXS data, 2) sedimentation velocity and sedimentation equilibrium AU experiments (see Chapter 4), or 3) dynamic light scattering measurements (Liu, et al.; unpublished). Since Klenow's Rg and I(0)/c concentration dependencies appear to be nonlinear and induced by high flux X-rays, Rg and I(0)/c values at infinite dilution were not calculated from these data sets.

Thus, the SSRL-Jul06 and SSRL-Jun07 SAXS curves for Taq and Klenow (see Table 3.2) were not used for modeling, instead the Taq and Klenow wide-angle scattering patterns were merged as described in Chapter 3 with the SAXS curves from SSRL-Jan03. Klentaq's WAXS curve was separately merged with both the SSRL-Jan03 and SSRL-Jun07 SAXS curves. All of the combined scattering curves were processed by GNOM, and shape models were created from each merged scattering curve using GASBOR as described in Chapter 3. GASBOR represents the scattering particle as a chain-like ensemble of small beads (dummy residues), each separated by 3.8 Å. This chain-like arrangement mimics the Cα backbone, and

the number of beads is equivalent to the total number of amino acid residues comprising the protein (the sequence is, however, arbitrary) [122].

5.2.2 GASBOR Modeling

GASBOR shape models of apo Klentaq, Klenow, and full-length Taq yielding the best fits to the measured X-ray scattering curves are shown in Figure 5.3. Each molecular envelope is docked with a corresponding atomic resolution structure as noted in the figure legend. Fits to the experimental data are excellent and do not significantly deviate between individual trials (see Figure 5.3 and Table 5.1). Bead model features were generally consistent among the individual trials, and crystallographic coordinates exhibited relatively good docking within each envelope. Individual models were then superimposed, averaged, and filtered as described in Chapter 3 to yield the final molecular envelopes shown in Figure 5.4.

Table 5.1. Parameters describing quality of GASBOR shape models.[a]

Polymerase	Goodness of Fit (χ)[b]		Normalized Spatial Discrepancy (NSD)[c]	
	Range	Average	Range	Average
Klentaq	1.32– 1.62	1.47	1.24 – 1.43	1.33
Klenow	1.04 – 1.10	1.07	1.15 – 1.39	1.26
Taq	1.28 – 1.49	1.40	1.61 – 1.86	1.77

[a] All values determined from a set of 12 models for each polymerase.

[b] $\chi = \left[\frac{1}{n-1} \sum_{j=1}^{n} \left(\frac{\langle c(q) \rangle I_{DR}(q_j) - c I_{\exp}(q_j)}{\sigma(q_j)} \right)^2 \right]^{0.5}$, where $I_{exp}(q)$ and $I_{DR}(q)$ are respectively the experimental and DR (dummy residue) scattering intensities at q_j, j = 1,….n, n is the total number of experimental data points being evaluated, c is a scaling factor, $\sigma(q)$ is the experimental data point error.

[c] Describes quality of the match between each bead model and its docked crystal structure. Each bead model was superimposed with corresponding atomic coordinates as shown in Figure 5.3.

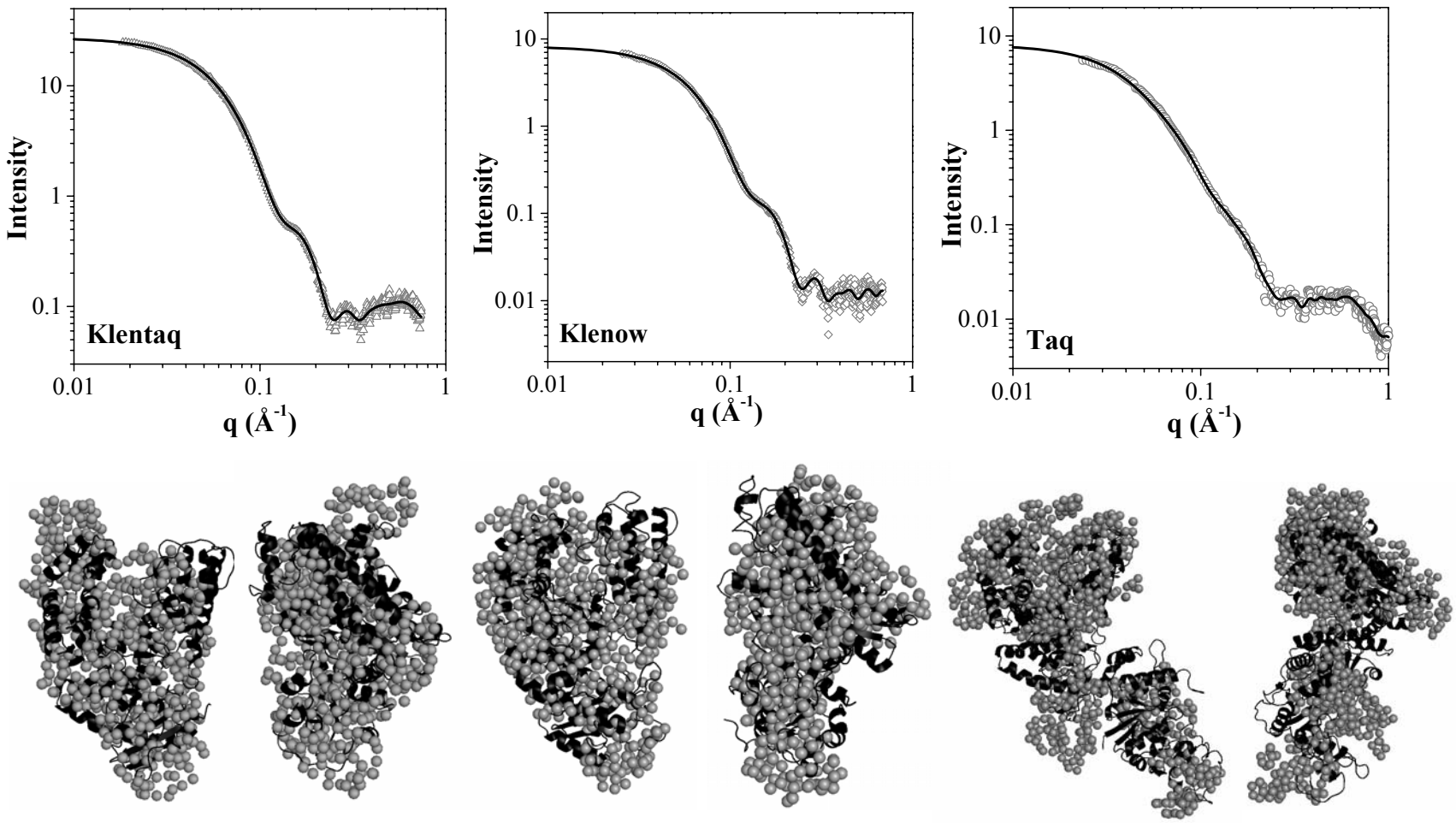


Figure 5.3. Best fit GASBOR *individual* shape models for Klentaq, Klenow, and full-length Taq DNA polymerases. *Top:* The log-log scale graphs contain the merged experimental X-ray scattering curves (open symbols) overlaid with the best fit scattering pattern (solid line) calculated from the respective GASBOR bead model depicted below the graph. *Bottom:* Corresponding bead models for each polymerase are shown in two different orientations below the respective scattering curves. The right model of each set is rotated -90° about the y-axis. Each bead model is docked with corresponding protein atomic coordinates from the following PDB entries: Klentaq fragment of 1TAQ (Klentaq), 1KRP [145] (Klenow), 1TAQ (full-length Taq).

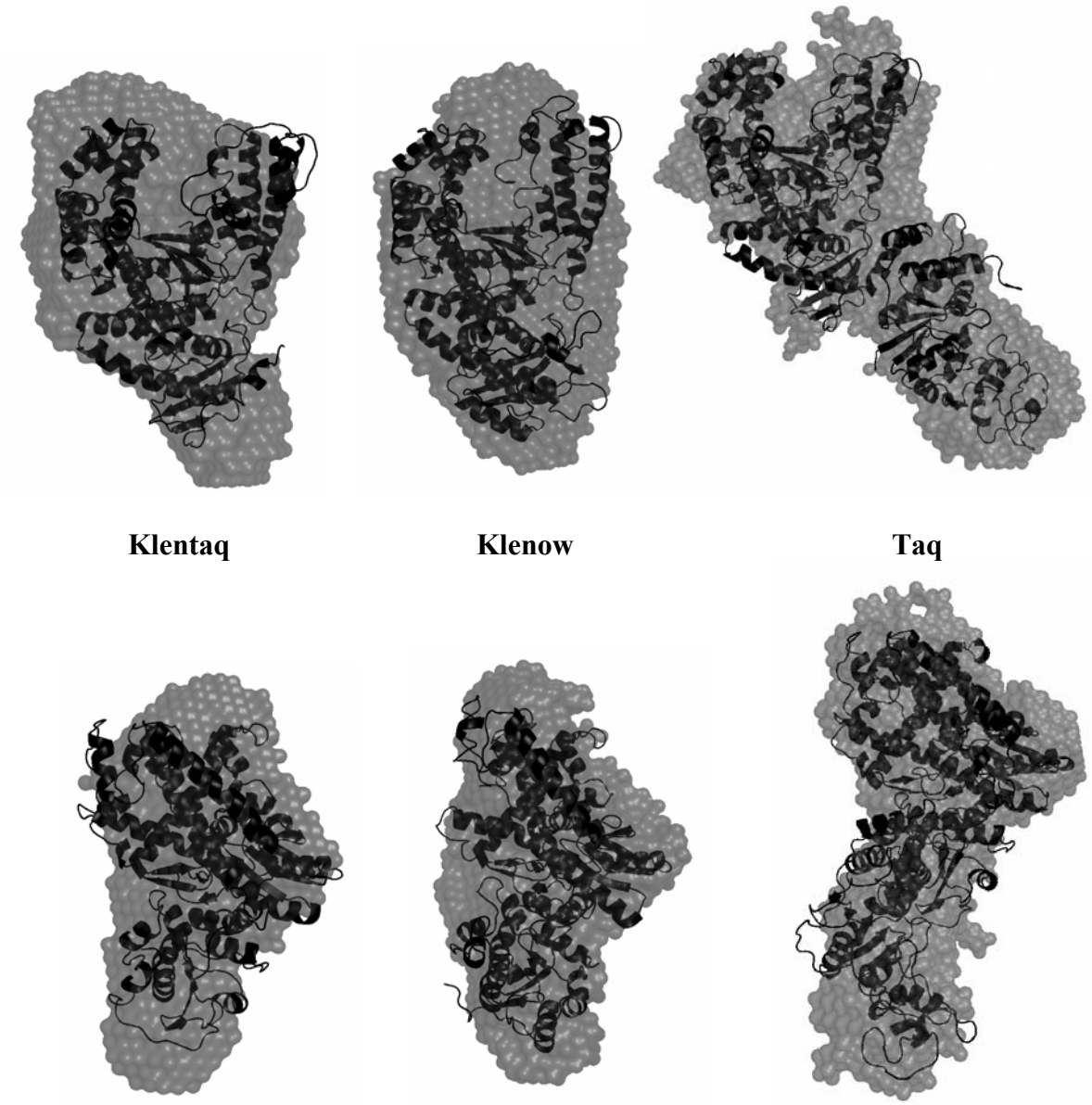


Figure 5.4. Averaged and filtered GASBOR shape models. *Bottom:* Models are rotated -90° about the y-axis relative to the top panel orientation. Molecular envelopes are docked with protein atomic coordinates Klentaq fragment of 1TAQ (Klentaq), 1KRP (Klenow), and 1TAQ (full-length Taq).

As reported in Table 5.1, the deviation in the goodness of fit parameter (χ) is less than 0.3 for each polymerase. The χ value is highly dependent upon the error of the data points, thus it is difficult to compare this goodness of fit parameter to similar values in the literature for other systems (most of which were collected on lower power beamlines and contain much more error, which can artificially decrease the χ value). For our data, however, which all have similar error ranges, the χ values indicate that the experimental data for Klentaq and Taq are fit equally well by the shape models, and that Klenow's scattering curve is the best fit of the three. Lower normalized spatial discrepancy (NSD) values indicate better agreement between the superimposed structures. NSD values for the individual bead models are somewhat higher than higher than for the averaged, filtered shape models of the polymerases.

Individual fit models, as shown in Figure 5.3, will vary from modeling run to modeling run on the same data. This is the result of the randomly seeded, initial arrangement of dummy residues, the iterative modeling process, and the fact that degenerate models will fit the scattering curve. These *individual* models are instructive for illustrating the highest resolution structures that fit the data. The scattering curve of the averaged, filtered model will not directly match the experimental data, whereas each *individual* model will. Because the solution set is degenerate, however, the averaged, filtered models, as shown in Figure 5.4, yield a molecular envelope representative of the most consistent *individual* model features that correspond to the global conformation of the macromolecule in solution.

5.2.3 Modeling of Taq DNA Polymerase in Solution Confirms Its Extended Conformation

Of the three polymerases, the Taq shape models demonstrated the most variation in individual bead model features. However, an overall extended, multi-domain conformation and good superimposition with the elongated Taq crystal structure were always maintained. The

averaging and filtering process smoothed out or eliminated many variable features of the individual Taq bead models (such as the protrusion between the proofreading and 5' nuclease domain visible in Figure 5.3), and yielded a final shape model that is strikingly similar to the elongated atomic resolution structure of Taq DNA polymerase (see Figure 5.4). This detailed *ab initio* shape model even reveals a hint of the cleft between the fingers and thumb subdomains. While this cleft is often also visible in the individual large fragment bead models, its position is not a stable feature and it is filled in as a result of averaging.

The fact that these shape models are obtained *ab initio*, without any input of crystal structure information, bears reemphasis. The resultant similarity between the solution shape envelopes and the known crystal structures lends strong validation to the accuracy of the derived structures, and supports the credibility of scattering derived structures of macromolecular species for which there exist no atomic resolution structures, such as those examined in the next chapter.

5.2.4 X-ray Scattering Shape Models Reveal Regions of Flexibility

While the crystallographic coordinates corresponding to Klentaq, Klenow, and the elongated conformation of Taq all superimposed extremely well with the respective final X-ray scattering shape models, visual inspection of Figure 5.4 reveals two notable features of the large fragment molecular envelopes that are not well correlated with the static atomic resolution structures and likely reflect regions of flexibility within these polymerases in solution. The shape models of Klentaq, and to a lesser extent Klenow, contain regions of excess volume both 1) above the fingers subdomain and 2) below the 3' proofreading domain.

5.2.4.1 Excess Volume and Conformational Flexibility

As shown in Figures 5.3 and 5.4, empty space both above the fingers subdomain and below the 3' proofreading domain of Klentaq is clearly observed in both the best fit *individual* bead model and the final averaged, filtered molecular envelope. Unlike the random protrusions observed in the *individual* Taq shape models, these two features of Klentaq's envelope were consistently observed and similarly positioned in the majority of its *individual* shape models, which explains why they were retained in the final averaged molecular envelope.

It should be noted that the Klentaq bead models presented herein were generated from combined SAXS and WAXS data collected on the same high flux beamline, whereas the merged scattering curves for full-length Taq and Klenow were combined from two different beamlines. This difference, however, cannot explain the unoccupied volume observed in the Klentaq shape models because nearly identical features were also present in the final molecular envelope obtained from a Klentaq scattering curve that was comprised of data from the two different beamlines and prepared in a manner similar to the merged X-ray scattering curves of Klenow and Taq (data not shown).

5.2.4.2 Proofreading Domain: N-terminus

The presence of wide angle X-ray scattering (WAXS) data makes the fit of simulated scattering from crystallographic coordinates particularly sensitive to missing or flexible regions of the macromolecule. For example, when simulated scattering curves from two Klenow crystal structures (protein coordinates only from 1KFD and 1KRP PDB entries), which are similar in overall conformation but differing in the total number of missing residues, are compared, the structure missing the most residues clearly does not fit the experimental scattering curve as well

(data not shown). Such gaps in the crystal structure can also appear as excess volume in the SAXS-WAXS derived structures.

As shown in Figure 5.5, we propose that the excess volume below the proofreading domain is nicely filled by a flexible N-terminus. Although the Klentaq fragment used to obtain the apo Klentaq crystal structure (1KTQ, as well as several other Klentaq + substrate co-crystal structures) actually consists of 559 amino acid residues, only the C-terminal 543 residues or less were reported due to disorder within the first 16 or more residues of the N-terminus [33, 39].

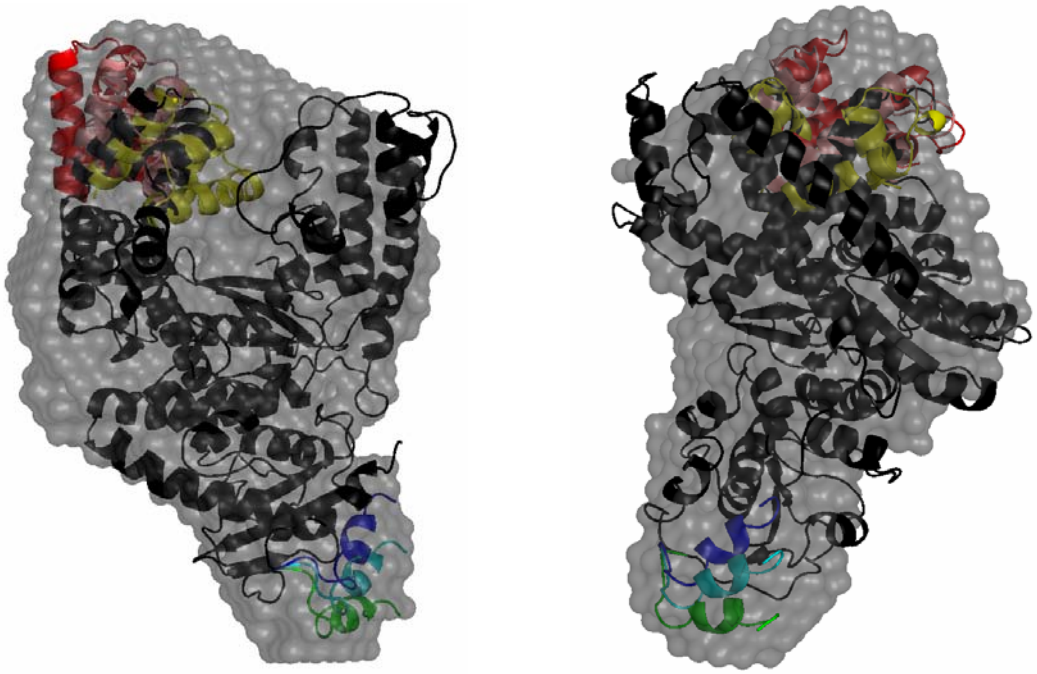


Figure 5.5. A flexible N-terminus and a flexible fingers subdomain can account for the excess volume of Klentaq's molecular envelope (gray). The atomic coordinates from the 1TAQ PDB entry corresponding to the Klentaq fragment (black, cartoon representation) are superimposed with the shape model. The N-terminus (residues 279 – 293) is shown in blue, cyan, and green (each color represents a separate conformation of the segment). Three different conformers of the fingers subdomain (residues 637 – 700) are depicted in red, pink, and yellow. Red and pink represent manually positioned fingers subdomain, whereas the yellow fingers subdomain is from the closed ternary complex of Klentaq bound to DNA and nucleotide (PDB ID: 3KTQ), and thus represents a documented motion of this domain. The different conformers shown in this figure were generated via rigid-body motions as described in the text.

The Klentaq fragment used in our X-ray scattering experiments is, however, comprised of 554 amino acids, and thus, has an 11 amino acid N-terminal extension compared to the 1KTQ crystal structure [91]. Therefore, to represent the full structure of our apo Klentaq, we extracted the atomic coordinates that correspond to the total number of residues comprising our Klentaq fragment (residues 279 – 832) from the 1TAQ PDB entry for full-length Taq, and found that these additional 11 amino acids can readily account for the excess volume in the proofreading domain. To generate the N-terminus conformers depicted in Figure 5.5, first, the loop between G286 and A293 was manually manipulated using PyMOL and oriented within Klentaq's molecular envelope as represented by the green residues. Then to fill in the remaining volume, the cyan and blue segments were positioned via rigid body rotation about the Cα of A293.

The proposal of a flexible N-terminus correlates well with the fact that the N-terminal residues are often disordered in crystallographic studies. According to the Klentaq shape model presented in Figure 5.5, this N-terminus, which joins the large fragment to the 5' nuclease domain in the full-length polymerase, does not appear to sample the conformational space occupied by these residues in the compact Taq, 1CMW, crystal structure (see Figure 4.1). It is notable, however, that the N-terminus does exhibit a range of motion, as shown in Figure 5.5, which is extremely similar to the orientation of the 5' nuclease domain in the extended conformation of full-length Taq DNA polymerase.

5.2.4.3 Fingers Subdomain

To account for the obvious region of excess volume above the fingers subdomain that is unoccupied by the static Klentaq crystal structure, we propose that flexibility within the tip of the fingers yields an equilibrium distribution of states in which the fingers can sample a wide range of conformational space as shown in Figure 5.5. The X-ray scattering curve for a solution of

conformationally heterogeneous particles is a weighted average of all conformational states, in which larger particles exert more influence on the final curve. Upon binding DNA and dNTP, the tip of the fingers subdomain (residues 637 – 700) consisting of helicies N, O, O1, and O2, undergoes rather large rigid body motions to yield a "closed conformation" ternary complex that is productive for nucleotide incorporation [39]. The motion of the fingers in the open to closed transition is shown in Figure 5.5 and can be broken down into two separate rotations, the largest of which is 40° [39]. Since the fingers move towards the thumb (see Figure 5.5), thus increasing unoccupied space above the fingers, it is difficult for the experimentally determined molecular envelope to indicate whether or not the crystallographically identified "closed conformation" [39] represents one of the suggested solution equilibrium states in apo Klentaq. This conformational change does, however, illustrate that the fingers are capable of rather large rigid body motions.

No extended conformational state for the tip of the fingers subdomain beyond the open configuration [39] has been previously reported, but the orientations colored red and pink in Figure 5.5 illustrate a possible range of motion of the fingers subdomain beyond that of the "open conformation" that is consistent with the experimentally determined molecular envelope. The tip of the fingers is connected to the base of this subdomain by two short linkers (marked by asterisks in Figure 5.6) joining helicies M to N and O2 to P. To orient the tip of the fingers as demonstrated by the red and pink domains in Figure 5.5, a hinge motion centered at Cα atom of residues of D637 and F700 is postulated.

The shape models of Klentaq, Klenow, and full-length Taq suggest that these polymerases do not share an equivalent degree of flexibility within their respective fingers subdomains. In Figure 5.4 an area of excess volume near the "top" of the Klenow's molecular

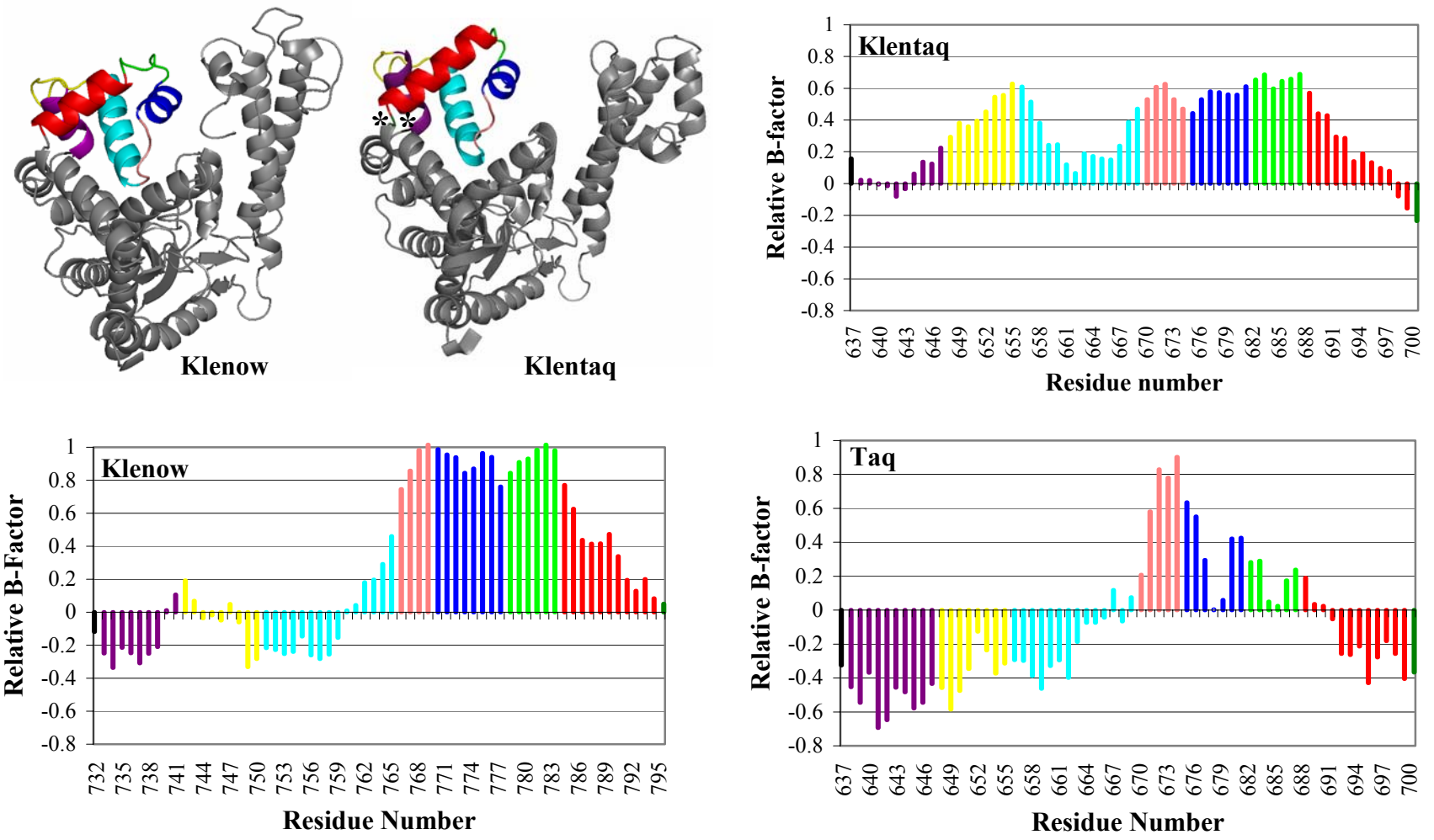


Figure 5.6. Normalized crystallographic B-factors indicate different degrees of flexibility within the fingers subdomains of Klentaq, Taq, and Klenow DNA polymerases. *Top left:* The polymerase domains of Klenow and Klentaq are shown in cartoon representation. The N, O, O1, and O2 helicies within the tip of the fingers subdomain are respectively colored purple, cyan, blue, and red, and the interconnecting linkers are colored yellow (N-O linker), pink (O-O1), and green (O1-O2). Matching colors are used to represent these secondary structural features within the bar graphs. Residue numbers correspond to full-length Taq for Taq and Klentaq and full-length *E. coli* Pol I for Klenow.

envelope is also visible, but when comparing Klentaq and full-length Taq, it is odd that only Klentaq exhibits apparent flexibility within the fingers subdomain, while full-length Taq does not. However, the results of a comparative analysis of the B-factors for this subdomain from the Klentaq, Klenow, and full-length Taq crystal structures (PDB IDs: 1KTQ [33], 1KRP [145], and 1TAQ [30]) shown in Figure 5.6 reveal that more residues within the tip of Klentaq fingers have a higher than average B-factor. The crystallographic B-factor, also known as the temperature factor, is one of the parameters used to refine the position of atoms within the crystal structure. High localized B-factors can indicate flexibility or disorder within that region of the molecule [146]. B-factors corresponding to the Cα atom of each residue within the polymerase domain (amino acids 420 – 832 for Klentaq/Taq and 516 – 928 for Klenow) were averaged for each protein. Then, each B-factor was normalized by the respective average. The relative B-factors in Figure 5.6 describe the deviation of each B-factor from the normalized average.

Although some fingers residues within full-length Taq and Klenow have positive relative B-factors that are higher than those of Klentaq, they only comprise a subset of the region of interest, and the remaining residues have a low or even negative relative B-factor. For Klentaq, the majority of the amino acids within the tip of the fingers subdomain have a higher than average B-factor. Therefore, collectively the fingers subdomain of Klentaq is more flexible than the respective regions of Klenow or Taq. Note in particular between Klentaq and Taq the complete change from positive to negative relative B-factors for the O-helix (cyan), the N-O linker (yellow), and the O2-helix (red) and significant decreases in relative B-factors in the O1 helix (blue) and O1-O2 linker (green).

The crystallographic B-factors correlate with the observed shape model differences and proposed local mobility within the polymerases, and both suggest that the presence of the 5’

nuclease domain in full-length Taq restricts the flexibility of the fingers subdomain. It is easy to imagine the fingers subdomain opening and closing in the manner illustrated in Figure 5.5 as it captures a nucleotide triphosphate from the solution and delivers it to the polymerase active site. The molecular envelopes generated from the X-ray scattering curves thus suggest that the range of motion of Klentaq's fingers subdomain in solution is larger than indicated by crystallographic data. Further experiments are needed to determine whether this flexibility persists upon DNA or dNTP binding.

Certainly, a much more detailed picture of the macromolecule in solution is obtained from shape modeling versus purely obtaining P(r) functions and associated scattering parameters (Rg and Dmax). Shape modeling is especially important in characterizing molecules for which no high resolution structures exist or those which deviate from crystallographic data. In the next chapter, we discuss shape models of polymerase-DNA complexes containing longer DNA substrates than have been crystallographically studied in any protein-DNA system.

CHAPTER 6

GLOBAL SOLUTION CONFORMATIONS OF KLENOW AND KLENTAQ IN COMPLEX WITH DNA: AN X-RAY SCATTERING CHARACTERIZATION

6.1 Introduction

Both a "polymerization" and an "editing" DNA binding mode have been characterized for Type I DNA polymerases based primarily on three-dimensional crystal structures of Klenow and Klentaq in complex with DNA substrates [37] and comprehensive biochemical characterizations of DNA binding by Klenow [12]. However, no three-dimensional structural data exists for the same polymerase binding DNA in both modes. Co-crystal structures for Klentaq + DNA bound in the polymerization mode and Klenow + DNA in the editing mode are shown in Figure 6.1. In the polymerase mode, the DNA is bound solely within the polymerase domain, while in the editing mode the duplex portion of DNA is bound in a cleft between the polymerase and proofreading domains with the 3' terminus of the single stranded region bound within the proofreading domain. A shuttling mechanism by which the 3' terminus of the DNA substrate is capable of translocation between the two active sites without the duplex portion of the DNA dissociating from the enzyme has been proposed to explain how Klenow and other similar proofreading DNA polymerases catalyze both 5' – 3' nucleotide incorporation and 3' – 5' nucleotide excision within two distinct active sites that are separated by ~35 Å [35, 37, 38]. Biochemical, kinetic, and mutational studies support this model in which the DNA substrate moves from one active site to the other without dissociating from the enzyme [52, 147, 148].

In the crystal structures for the "polymerization" versus the "editing" mode, it is evident that the DNA protrudes from the respective complexes in different orientations, dependent upon the binding mode. Furthermore, DNA binding experiments performed in our laboratory

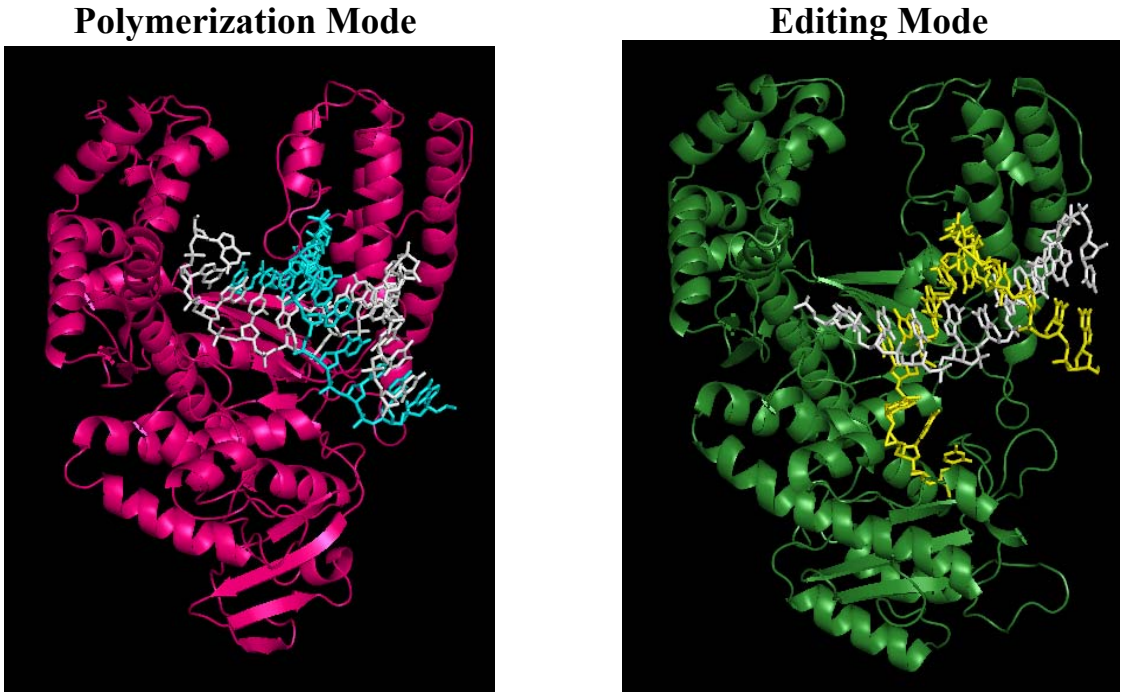


Figure 6.1. Polymerization and editing DNA binding modes for Klentaq and Klenow. *Left:* Klentaq bound to DNA in the polymerization mode. *Right:* Klenow complexed with DNA in the editing mode. The single-stranded 3' overhang interacts with the proofreading domain, such that the terminal nucleotide is positioned near the 3' exonuclease active site. The DNA template strand in both structures is colored white, while the DNA primer strands are colored cyan and yellow, respectively, in the 4KTQ and 1KLN structures.

demonstrate that DNA binding to Klentaq and Klenow is structure dependent and polymerase specific. Klenow exhibits higher affinity for primed-template DNA over blunt-ended duplex DNA, while Klentaq demonstrates identical binding affinity for both DNA structures (Wowor and LiCata, unpublished). These thermodynamic results suggest the existence of additional binding mode differences between the two polymerases. In this chapter a combination of small and wide angle X-ray scattering was used to characterize the global solution conformations of Klentaq and Klenow in complex with both primed-template and blunt-ended DNA duplexes. Comparisons against atomic coordinate models of these polymerases interacting with DNA in

both the polymerization and editing modes suggest that Klentaq binds both DNA structures in the polymerization mode, while Klenow binds both substrates in the editing mode. The exact details of these interactions in solution seem to differ slightly from the crystallographically-derived atomic coordinate models and require further investigation.

6.2 Results and Discussion

The polymerase-DNA complexes examined in this X-ray scattering study were designed so that both the protein and DNA components would contribute to the overall shape of the complex in solution. As such, the DNA substrate must be long enough to significantly protrude out of the polymerase envelope. Previous X-ray scattering measurements of Klenow bound to a primed-template (pt)-13/20mer, which is slightly longer than the DNA constructs in the polymerase-DNA co-crystal structures, indicate that this DNA substrate is virtually invisible with respect to the protein shape scattering due to its relative size [95]. Thus, the global solution conformations of Klentaq and Klenow each bound to a pt-63/70mer and a blunt-ended double stranded (ds)-63/63mer were examined. In addition to fulfilling the length criteria, pt-63/70mer and ds-63/63mer were chosen as the DNA substrates for these X-ray scattering experiments because the binding thermodynamics of these specific DNA constructs to Klentaq and Klenow have been previously characterized [81, 94, 95] (Wowor and LiCata, unpublished).

6.2.1 Construction of Atomic Coordinate Models (ACMs)

In the absence of high resolution structures that are directly comparable to the experimental polymerase-DNA complexes, atomic coordinate models (ACMs) were generated based on extrapolation of the known co-crystal structures and using simulated DNA constructs. These ACMs and their structure-based X-ray scattering parameters (calculated by CRYSOL analysis [121]) are used herein for comparative purposes against our experimental results. To

yield the projected polymerization and editing mode ACMs for our polymerase-DNA complexes in solution, simulated ds-63/63mer and pt-63/70mer DNA constructs were aligned, as illustrated in Figure 6.2, with the much shorter 4KTQ and 1KLN crystallographic DNA substrates. While superimposing the simulated DNA constructs with the crystallographic DNA substrates, we were careful to avoid introducing any arbitrary positional deviations among the structures that might translate into artifactual differences in the overall dimensions of the polymerization and editing mode complexes.

The crystallographic DNA substrates do not adopt a perfect B-form conformation like the simulated constructs. DNA bound to Klentaq is mostly B-DNA with the terminal three base pairs closest to the polymerase active site assuming an A-form conformation [39], while the duplex portion of the DNA bound to Klenow adopts a distorted B-DNA structure [37]. As a consequence of these conformational differences between the simulated and crystallographic DNA molecules and since A-DNA has a shorter rise/base pair, the best alignments were obtained by positioning the 3' terminal nucleotide of the simulated DNA primer strand slightly farther into the polymerase envelope and closer to the fingers subdomain by approximately 0.5 – 1 base pair (< 3.4 Å). Chapter 3 contains a more detailed description of the preparation of the ACMs.

6.2.2 Concentration Dependence of Scattering Parameters

SAXS curves from each DNA substrate and polymerase + DNA complex were measured at 4 concentrations between 20 and 80 μM to check for interparticle interference and/or X-ray induced aggregation. An additional measurement was taken for ptKTQ at 10 μM. SAXS curves from the ptKTQ concentration series are shown in Figure 6.3 and exhibit a concentration dependent pattern that is shared by the SAXS curves for the other polymerase + DNA complexes and the isolated DNA substrates. At very low scattering angles, corresponding to q values below

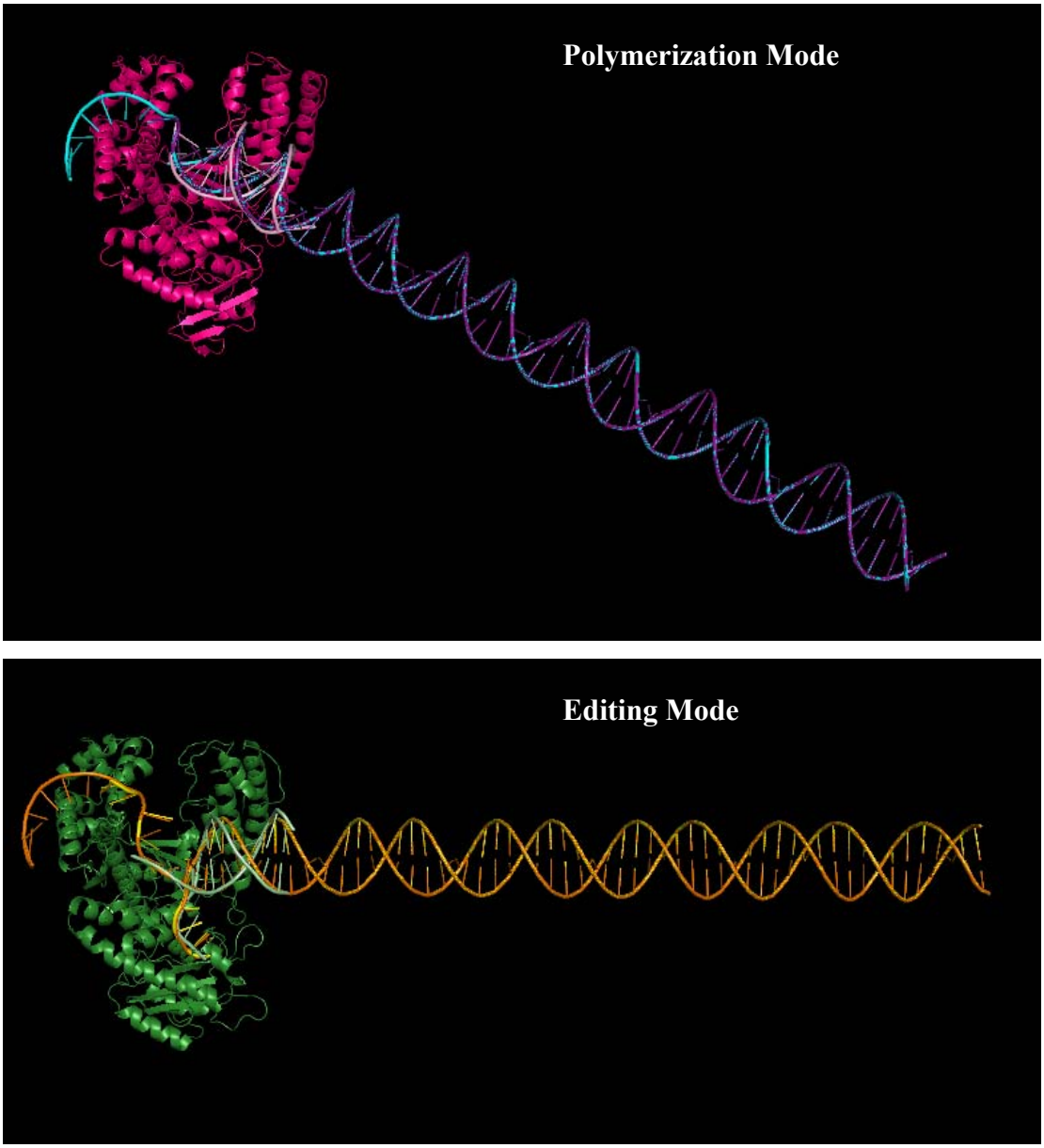


Figure 6.2. Preparation of atomic coordinate models (ACMs) via alignment of simulated pt and dsDNA constructs with crystallographic DNA substrates bound to KTQ *(top)* and KLN *(bottom)*, respectively, in polymerization and editing modes. Crystallographic coordinates are from 4KTQ and 1KLN PDB entries. B-DNA coordinates were simulated using the Nucleic Acid Builder program [136, 137] as described in Chapter 3. In the top panel, KTQ is shown with the crystallographic DNA (light pink), while the cyan and purple DNAs are the simulated ptDNA and dsDNA, respectively. All DNAs in the top panel are bound in the polymerization mode. In the bottom panel, KLN is shown with the crystallographic DNA (light green), while the orange and yellow DNAs are the simulated ptDNA and dsDNA, respectively. All DNAs in the bottom panel are bound in the editing mode. The simulated pt and dsDNA structures are perfectly superimposed within their duplex regions for each binding mode, and thus the DNAs appear speckled with the colors of the individual DNA constructs.

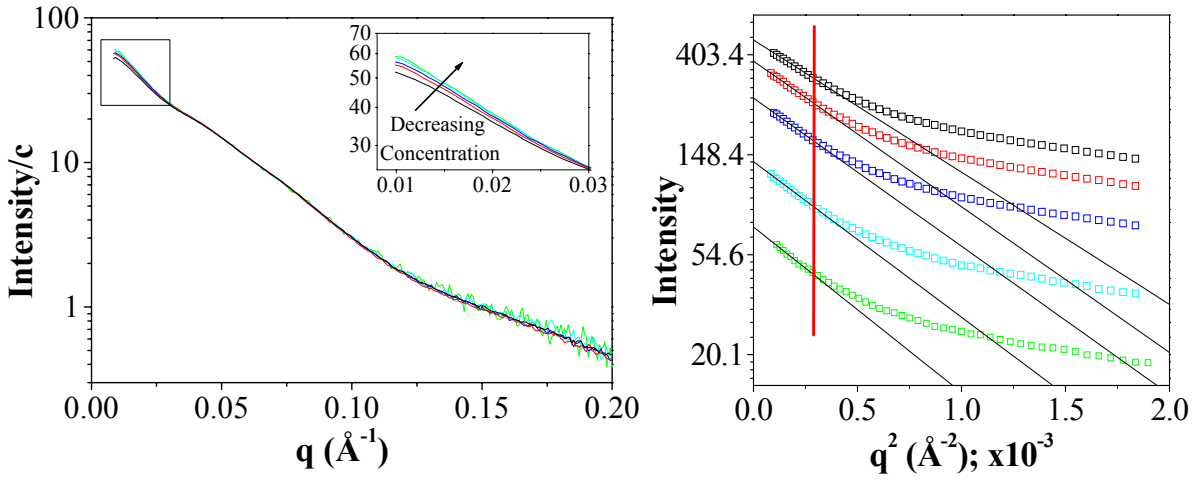


Figure 6.3. Concentration dependent influence of interparticle interference on the SAXS curves of binary complexes and isolated DNA. Only scattering curves collected from the ptKTQ concentration series are shown, but the other binary complexes and isolated DNA molecules exhibit similar concentration dependent patterns due to repulsive interparticle interactions. *Left:* Concentration normalized scattering intensity plotted on a $\log_{10}$ scale as a function of q. The range of q values affected by concentration dependent phenomena is restricted to very small scattering angles (boxed region), and is particle specific. The boxed region is magnified in the inset graph. *Right:* Corresponding Guinier plots of ptKTQ concentration series. The ordinate axis contains un-normalized scattering intensities plotted on a natural log scale. Low q data points clearly influenced by parasitic scattering near the beamstop are not plotted and were excluded from analysis. Solid black lines are the linear Guinier fits to the scattering data up to the vertical red line and extrapolated to the abscissa. Concentrations are 10 μM (green), 20 μM (cyan), 40 μM (blue), 60 μM (red), and 80 μM (black).

0.05 $Å^{-1}$, I(q) increases with decreasing concentration, resulting in a classic concentration dependence pattern characteristic of interparticle interference. This common experimental phenomenon underscores the need to routinely collect scattering curves at different sample concentrations. In the case of these isolated DNAs and protein-DNA complexes, the interparticle interference is likely due to repulsive electrostatic interactions between the rather large, negatively charged DNA molecules used in this experiment, and this result is not surprising since DNA is known to exhibit repulsive electrostatic interparticle interference at high DNA and low salt concentrations [149-151]. It is noteworthy that the scattering curves of apo Klentaq and Klenow do not show similar signs of repulsive interactions (see Figure 5.1).

The influence of interparticle interference on the experimentally measured scattering curve, $I_{exp}(c,q)/c$, is expressed by the structure factor, S(c, q), in equation 6.1.

$$\frac{I_{\exp}(c,q)}{c} = P(q) * S(c,q) \qquad \text{Equation 6.1}$$

P(q) is the form factor, which describes the size and shape of the macromolecule. The structure factor represents long range interactions between individual macromolecules and is dependent upon the solution environment (solvent composition, temperature, pH, etc) [152-155]. At zero concentration, the structure factor, S(c, q), equals 1 [124] and, according to equation 6.1, only the form factor describes the shape of the scattering curve. It is not currently possible to computationally or theoretically predict if or how a structure factor may influence the measured scattering curve, and thus its presence must be experimentally determined, as is done herein, by measuring a macromolecule's scattering curve at several different concentrations.

The structure factor significantly affects only the very low q scattering intensities. As demonstrated in Figure 6.3, the ptKTQ scattering curves all overlap above 0.05 A^{-1}. However, it is this low q region that describes the global size and shape of the macromolecule and thus, significantly impacts the radius of gyration and the forward scattering intensity. Concentration dependencies for the Rg and I(0) values determined for each macromolecule using the indirect, inverse Fourier transformation method implemented by GNOM [119] are shown in Figure 6.4. Parallel analyses using Guinier plots show the same patterns (data not shown). Since the structure factor does not influence the scattering data at zero concentration, linear extrapolation of I(0)/c(c) and Rg(c) to infinite dilution (see Figure 6.4) yields $I(0)/c_{c=0}$ and $Rg_{c=0}$ values that correspond solely to the particle size and shape.

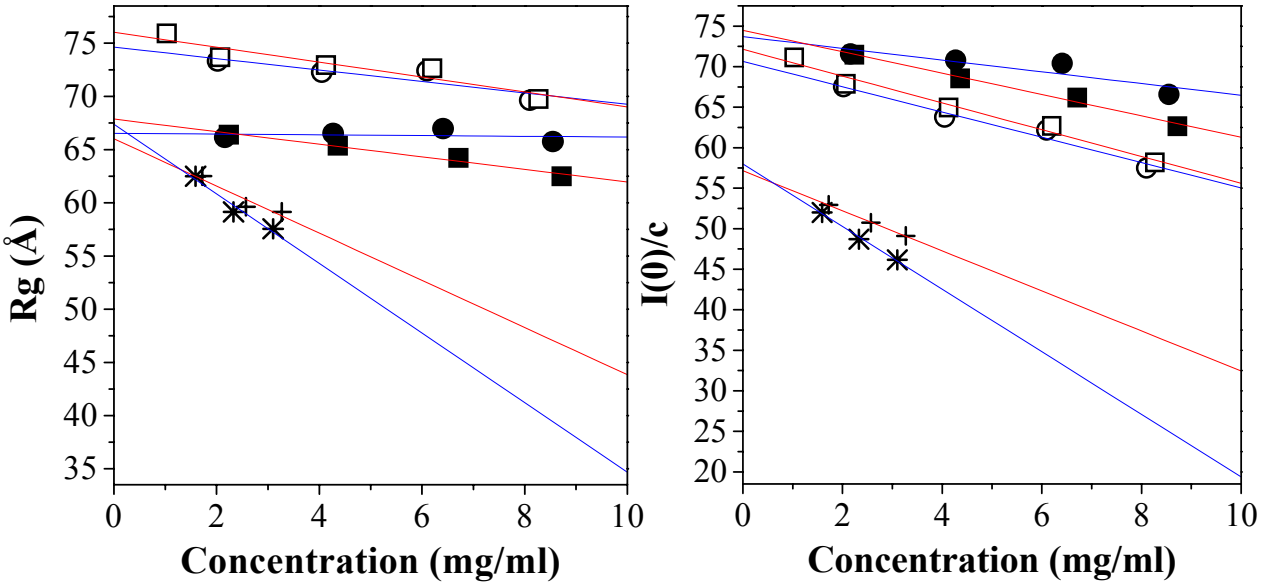


Figure 6.4. Concentration dependence of GNOM derived Rg and I(0). For these graphs, the molar concentrations (10 – 80 μM) were converted to mass concentrations. Concentration dependencies are shown for ptDNA (+), dsDNA (✳), ptKTQ (□), dsKTQ (○), ptKLN (■), and dsKLN (●). Errors on Rg and I(0)/c are smaller than the symbols. Solid lines are the linear fits to each data set: ptDNA (red) and dsDNA (blue). The linear fits of the polymerase + DNA complexes are color coded according to the DNA component. Extrapolation of these fits to infinite dilution yields the Rg and I(0)/c values reported in Tables 6.1 and 6.2 .

For each DNA substrate and polymerase + DNA complex, both Rg and I(0)/c decrease with increasing concentration in a linear fashion (see Figure 6.4). Rg and I(0)/c for the same macromolecule exhibit similar concentration dependent patterns. Interestingly, the ds-63/63mer and pt-63/70mer DNA substrates do not demonstrate equivalent interparticle interactions; dsDNA scattering parameters seem to be more significantly influenced by a structure factor. It is possible that dsDNA, which is a perfectly matched blunt-ended negatively charged DNA molecule, can partially align in solution in a manner dictated by electrostatic repulsion to give rise to a larger structure factor. In contrast, the ptDNA, which has a seven nucleotide template overhang, may not be able to adopt a similar repulsive network due to the flexible single stranded region.

One focus of this study was to determine if the global solution conformations of Klentaq and Klenow DNA bound complexes are dependent upon nature of the polymerase and/or the structure of the DNA substrate. The binding of either pt or ds DNA to Klentaq yields complexes that are nearly equivalent in size and shape as a function of concentration as can be seen from their nearly overlapping lines in Figure 6.4. The structure factor affecting ptKLN seems very similar to that of ptKTQ and dsKTQ, as the Rg and I(0)/c concentration dependencies are nearly parallel for these three complexes. However, the slope of the concentration dependence of the dsKLN scattering parameters is reduced compared to the other polymerase + DNA complexes in Figure 6.4. Thus, while the solution conformations of the ptKLN and dsKLN seem to be very similar at ~ 2 mg/ml, at higher concentrations the global structures or interparticle interactions of ptKLN and dsKLN diverge. Since the molecular behavior that causes interparticle interference is poorly understood in general, it is difficult to explain the exact molecular cause of this high concentration difference for dsKLN, however, the observed concentration dependence of dsKLN is likely due to the presence of both attractive and repulsive interparticle interactions. Thus, at high concentrations dsKLN may begin to oligomerize or aggregate, whereas the stoichiometry of the other polymerase + DNA complexes seems to remain 1:1. Differences between the binding of pt and ds DNA to KLN, but not to KTQ, have also been observed on the thermodynamic level (Wowor and LiCata, unpublished).

6.2.3 Molecular Mass Determination and Associated Results

Experimental molecular masses shown in Table 6.1 were calculated from the normalized forward scattering intensity at infinite dilution by calibration against lysozyme as described in Chapter 3.

Table 6.1. Comparison of experimental and calculated molecular masses.

Complex or Isolated DNA	GNOM	Molecular Mass[b,c]		
	$I(0)/c_{c=0}$[a]	Calculated	Experimental	Δ
	a.u.[d]	*kDa*		%
ptKLN	74.5 ± 0.6	109.0	104.5 ± 0.8	-4.1
dsKLN	73.7 ± 1.5	106.9	104.8 ± 2.1	-1.9
ptKTQ	72.1 ± 0.7	103.3	98.7 ± 1.0	-4.5
dsKTQ	70.7 ± 1.1	101.2	97.9 ± 1.5	-3.3
ptDNA	57.2 ± 0.2	40.9	39.2 ± 0.1	-4.3
dsDNA	58.0 ± 0.8	38.8	39.7 ± 0.5	2.4

[a] $I(0)/c_{c=0}$ errors are the errors from the linear fits in Figure 6.4.
[b] Molecular masses were determined by calibration of $I(0)/c_{c=0}$ for each scattering particle against lysozyme's $I(0)c_{c=0}$ of 5.6.
[c] Molecular mass errors are $I(0)c_{c=0}$ error propagated through the calibration against lysozyme.
[d] a.u. = arbitrary units

According to equilibrium DNA binding studies with Klentaq and Klenow and the pt-63/70mer and ds-63/63mer [81, 94, 95] (Wowor and LiCata, unpublished), the concentrations used in these X-ray scattering experiments are all at least 250-fold higher than the Kd's of complex formation, and thus each sample is expected to contain nearly 100% complex with 1:1 stoichiometry. The calculated monomeric molecular mass for each isolated DNA was determined from its known chemical composition, and the expected mass of each complex was tabulated to represent a 1:1 polymerase:DNA complex (see Table 6.1). Comparison of the experimentally determined values with the calculated molecular weight yields direct information concerning the oligomeric state of the scattering particle and the sample monodispersity.

All but one of the experimentally determined molecular weights reported in Table 6.1 are smaller than the expected values by less than 5%. The one exception, the dsDNA construct, has an experimental molecular mass that is slightly higher, but still within 5% of its theoretical value. The extrapolated molecular weights indicate that the isolated DNA molecules are monomeric,

and that each complex forms with 1:1 stoichiometry. The similarity between the experimental and calculated molecular masses also indicates that the polymerase + DNA complex solutions are essentially monodisperse, i.e. the 1:1 complexes are not significantly mixed with unbound protein and DNA or complexes of other stoichiometries. Even if only ~1% of 2:1 or 1:2 polymerase:DNA binary complexes formed, the experimental molecular mass would be higher than the expected value for a 1:1 complex; however, all of the experimentally determined masses are slightly lower than the calculated values. On the other hand, if the polymerase and DNA did not interact, then the experimentally determined molecular weight would be a concentration weighted average of the individual molecular weights, and for these experiments would be approximately 50% lower than the expected value expected for a 1:1 complex.

X-ray scattering curves collected for the isolated DNA constructs, as well as for apo Klentaq and apo Klenow (which were discussed in Chapter 5) were used for comparison against the polymerase-DNA complex X-ray scattering data. Like the isolated DNA molecules, apo Klentaq is monomeric under the conditions of the X-ray scattering experiment; Klentaq's molecular weight is within 10% of the monomer value. Apo Klenow, on the other hand, aggregated when examined with the high flux X-ray beam (see Figures 5.1 and 5.2). However, under these same experimental conditions, when bound to ptDNA, Klenow does not exhibit X-ray induced aggregation (as demonstrated by Figure 6.4). Whether or not this is also the case for Klenow bound to dsDNA cannot be clearly ascertained due to the differences between dsKLN and the other polymerases in Figure 6.4 and Table 6.1 as discussed above.

6.2.4 X-ray Scattering Curve Transformations Reveal Shape Characteristics of the Global Conformations of the Complexes and Isolated DNAs

The DNA constructs used in this study were 63 – 70 nucleotides in length, but shorter than the persistence length of duplex DNA which is approximately 150 base pairs or ~500 Å

[156]. The persistence length is a macromolecular property that describes the rigidity of a polymer chain, and this length marks the transition between rod-like (length < persistence length) and random walk (length > persistence length) behavior of the macromolecule. Thus, our DNA constructs both in isolation and complexed with polymerase are expected to maintain a rod-like B-DNA conformation. The X-ray scattering curve transformations shown in Figures 6.5 and 6.6 display patterns classically associated with rigid rod-like particles for the pt and dsDNA constructs, and a combination of features characteristic of both rod-like and globular shapes for the DNA bound complexes.

Additional information concerning the cross-section of rod-like macromolecules can be acquired by multiplying the X-ray scattering curve (I(q)) by the scattering vector (q), which effectively removes the long-dimension data from the scattering curve [105, 125]. Figure 6.5 contains a Guinier plot of the transformed scattering curve. Rg_{xs} values, reported in Table 6.2, were calculated from the slope of the linear regime. Unlike the Rg values that were affected by interparticle interference and exhibited a concentration dependence, Rg_{xs} values were derived from the scattering curves at q vectors > 0.05 $Å^{-1}$ and were concentration independent.

Both isolated DNA curves depicted in Figure 6.5 exhibit an extensive linear range, which is characteristic of uniform, rigid rod-like particles [105], and the parallel nature of the linear fit indicates that the pt and dsDNA constructs share a similar cross-sectional radius of gyration (see Table 6.2). The polymerase-DNA complexes also display a region of linearity within the cross-sectional Guinier plot, albeit characterized by a more negative slope and less extensive q^2 range compared to that of the isolated DNAs. The flattening of the low q^2 scattering data for the complexes to the left of the vertical green line, shown in Figure 6.5, is a classic indication of a globular domain [105].

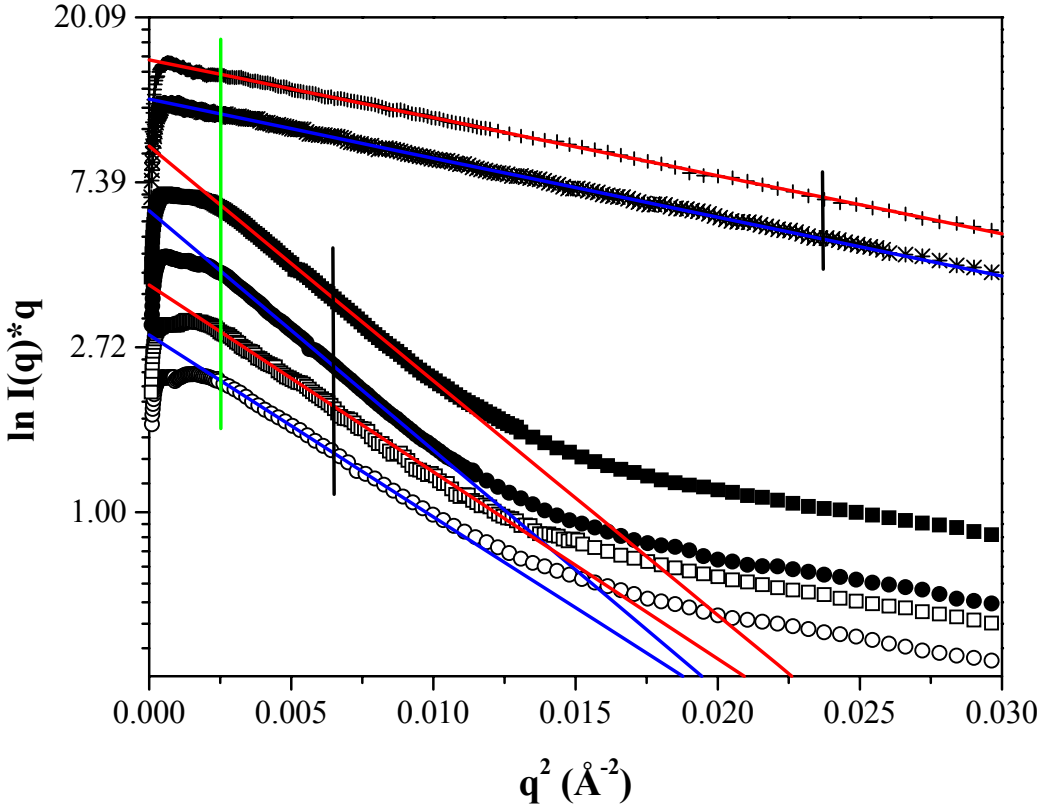


Figure 6.5. Determination of the cross-sectional radius of gyration (Rg_{xs}) by Guinier analysis. X-ray scattering curves (I(q)) for ptDNA (+), dsDNA (✳), ptKLN (■), dsKLN (●), ptKTQ (□), and dsKTQ (○) were multiplied by the scattering vector (q) to eliminate information concerning the long dimension of the scattering particles, and then graphed on a natural logarithmic scale as a function of q-squared. Rg_{xs} values reported in Table 6.2 were calculated from the slope of the linear region of each curve (similar to Guinier Rg determination) up to $q_{max}*Rg_{xs} \leq 1.3$. The boundaries of the linear ranges used for regression analysis are marked by the green and black vertical lines. Compared to the isolated DNA constructs, the polymerase-DNA complexes exhibit a reduced linear regime. Errors on data points are smaller than symbols. Solid lines are the linear fits to each data subset: ptDNA (red) and dsDNA (blue). The linear fits of the polymerase + DNA complexes are color coded according to the DNA component. Scattering curves are arbitrarily shifted along the y-axis for clarity.

Table 6.2. Experimentally determined and ACM-derived X-ray scattering parameters.

Complex or Isolated DNA	Experimental					Simulated[a]					
	GNOM		Rg_{xs}	D_{xs}[b]	L[c]	Rg		Dmax			
	$Rg_{c=0}$	Dmax				Pol	Edit	Pol	Edit	Rg_{xs}	D_{xs}[b]
	Å					Å					
ptKLN	67.9 ± 0.4	230	16.9 ± 0.1	47.8 ± 0.3	227.8 ± 1.5	68.7	68.8	243	236	16.3	46.1
dsKLN	66.5 ± 0.8	230	17.0 ± 0.1	48.0 ± 0.4	222.7 ± 3.0	68.8	68.6	243	236	15.8	44.7
ptKTQ	76.0 ± 0.8	260	15.1 ± 0.1	42.7 ± 0.3	258.0 ± 2.9	69.9	69.2	244	237	15.1	42.7
dsKTQ	74.6 ± 1.1	255	15.3 ± 0.2	43.3 ± 0.5	252.9 ± 4.1	69.8	69.0	244	236	14.7	41.6
ptDNA	66.0 ± 2.0	240	8.4 ± 0.1	23.8 ± 0.2	226.8 ± 7.0	61.6 - 65.9		217 - 241		6.9	19.5
dsDNA	67.4 ± 1.7	230	8.5 ± 0.2	24.0 ± 0.6	231.6 ± 6.0	61.6		217		6.9	19.5

[a] Simulated scattering parameters were determined from the simulated X-ray scattering curves calculated from the ACMs using CRYSOL [121].

[b] $D_{xs} = 2Rg_{xs}\sqrt{2}$ [157]

[c] $L = \sqrt{(Rg^2 - Rg_{xs}{}^2) * 12}$ [105]

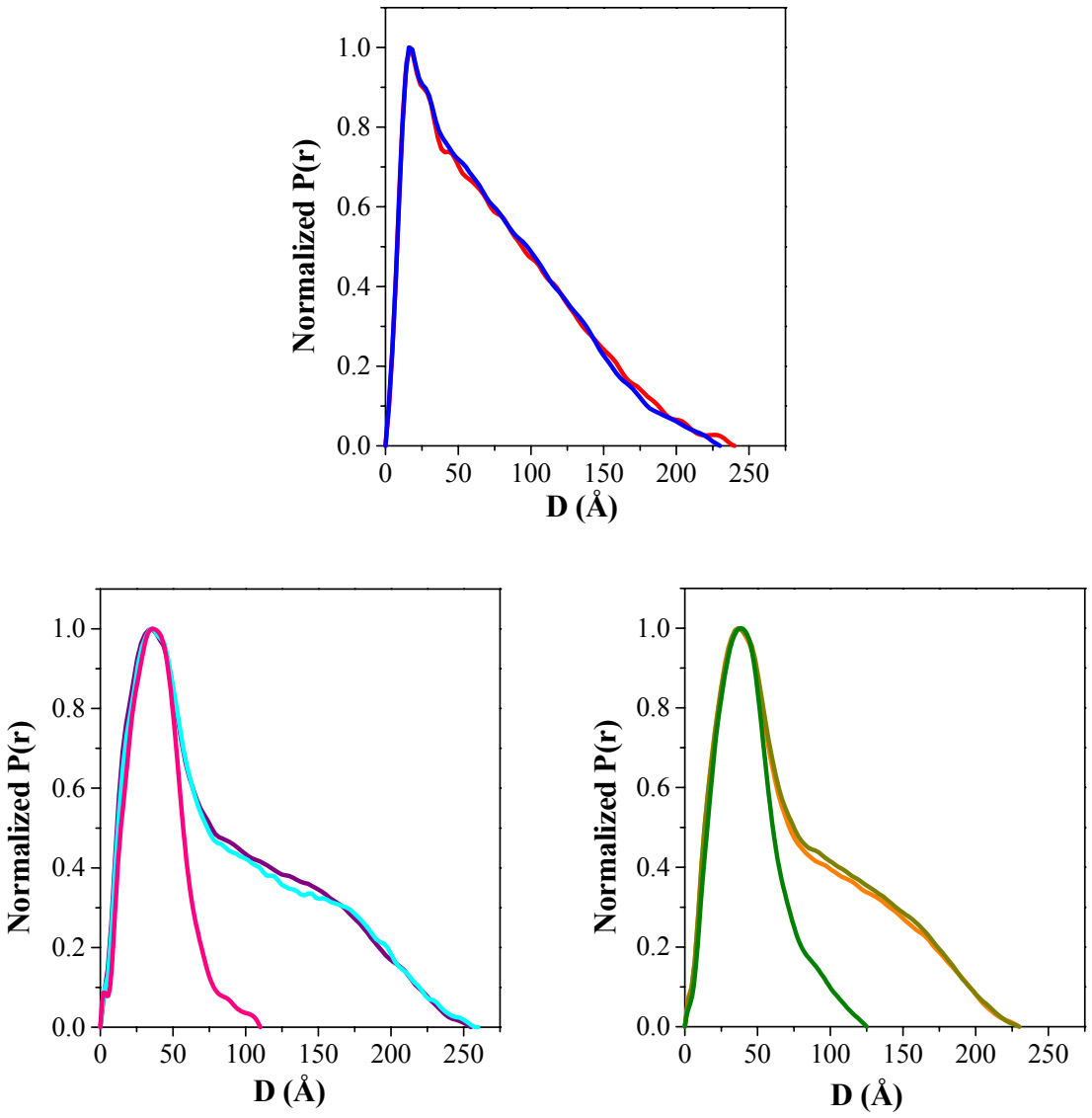


Figure 6.6. P(r) (pair distance distribution functions) yield size and shape information about the polymerase + DNA complexes and isolated protein and DNA components. All P(r) transformations were computed by GNOM from the X-ray scattering curves and then normalized to 1. *Top:* pt-63/70mer (red) and ds-63/63mer (blue) isolated DNA constructs display P(r) functions that are quite different from the apo polymerases and complexes. *Bottom:* Polymerase + DNA complexes, ptKTQ (cyan), dsKTQ (purple), ptKLN (orange), and dsKLN (yellow), are plotted together with the respective apo polymerase, apoKTQ (pink) or apoKLN (green), showing that the complexes retain the peak and some of the P(r) curve shape corresponding to the globular protein component.

The pair distance distribution functions or P(r) functions shown in Figure 6.6 were calculated from the merged experimental SAXS/WAXS curves (prepared as described in Chapter 3) using the program GNOM [119]. For the isolated DNA constructs (top panel), the extremely asymmetric distribution of scattering pair distances in the P(r) function and the nearly linear decay as a function of distance from the peak to the Dmax are both classic indications of a uniform rod-like particle shape.

The polymerase-DNA complexes (see bottom panels of Figure 6.6) also display an asymmetric P(r) function, characteristic of an elongated particle shape. However, the peak position and shape specifically correspond to that of the globular protein component comprising the complex. Although the isolated DNAs, KTQ complexes, and KLN complexes differ from one another either in particle shape or Dmax, the significant structural similarities between 1) the isolated pt and dsDNA constructs, 2) the two DNA bound KTQ complexes, and 3) the two DNA bound KLN complexes are readily apparent from these P(r) functions.

6.2.5 The Experimentally Determined Dimensions of the pt-63/70mer and ds-63/63mer in Solution Are Larger than Predicted by the Simulated pt and dsDNA Constructs

Experimentally, the pt-63/70mer and ds-63/63mer share nearly identical global conformations and associated X-ray scattering parameters (see Table 6.2). However, the measured dimensions of isolated DNA molecules in solution are larger than the respective values calculated from the simulated B-DNA ACMs. Since the diameter of B-form DNA is a characteristic dimension used to describe DNA structure, each experimental and calculated Rg_{xs} value was converted to a cross-sectional diameter (D_{xs}) using equation 6.2 [157], which relates these two parameters for rod-shaped macromolecules (see Table 6.2).

$$D_{xs} = 2Rg_{xs}\sqrt{2} \qquad \text{Equation 6.2}$$

The diameter of isolated double-stranded B-DNA is approximately 20 Å [158, 159] as reflected by our simulated DNA constructs (see Table 6.2). The experimentally determined cross-sectional diameter of the isolated DNA molecules in solution is ~4 Å wider than expected for B-DNA (see Table 6.2). The increased width of the experimental DNA most likely reflects the layer of counterions surrounding the DNA. This concentrated counterion layer would have an electron density contrast that is higher than that of the bulk solution, thus making the DNA itself appear slightly larger than expected. Hydrodynamic measurements of the rotational and translational frictional coefficients of DNA in solution, likewise, yield cross-sectional diameters of 22 – 26 Å [160]. This phenomenon, which is not fully accounted for in the scattering curves generated by the program CRYSOL, is similar to the increased contrast of the hydration layer surrounding proteins, but is more pronounced due to the significant counterion layer surrounding DNA.

In addition to the increased width, the DNA constructs also exhibit larger than expected Rg and Dmax values (see Table 6.2). These additional enlarged dimensions may also be explained by the layer of counterions that are detectable by X-ray scattering. The measured Rg and Dmax values are, respectively, 6 and 13 Å larger than the structure-based values for dsDNA, while for the ptDNA construct they are near the maximum of the calculated range. The simulated pt-63/70mer ACM was originally modeled with a rigid helical ss 5' template overhang (as shown in Figure 3.11), whereas in solution this ss overhang would surely be quite floppy and possibly not even resolved by X-ray scattering. Thus, the calculated Rg and Dmax values for the ptDNA construct are represented by a range of values in Table 6.2.

6.2.6 KTQ + DNA Complexes Exhibit Larger Global Dimensions than KLN + DNA Complexes

Experimentally, the global structural parameters (Rg, Dmax, etc.) of DNA bound KTQ are not altered by the DNA substrate structure (pt versus ds), beyond expected subtle differences likely due to the flexible ss overhang of the ptDNA construct. The same is separately true for the two DNA bound KLN complexes. The global structure of the polymerase-DNA complexes does, however, differ between the two polymerases. Complexes of KLN + either DNA substrate exhibit significantly smaller Rg and Dmax values than do the KTQ + DNA complexes. Compared to the KLN + DNA complexes, the DNA bound KTQ complexes are approximately 8 Å larger in Rg and 30 Å longer in Dmax (see Table 6.2). A difference of ~30 Å in length corresponds to an ~8 base pair difference in the length of protruding B-form DNA that must be accounted for in the KTQ versus KLN complexes.

Since selection of the Dmax during the iterative GNOM analysis of the SAXS data is somewhat subjective and since Rg typically increases with increasing Dmax, Rg's for KTQ and KLN complexes were also computationally compared at the same fixed Dmax. However, even under these highly forced constraints, the KTQ complexes were always larger than the KLN complexes by at least 5 Å. Guinier analysis also reveals similar Rg differences between KTQ and KLN DNA bound complexes. Thus, the global conformations of KTQ and KLN DNA bound complexes are polymerase-specific, and KTQ + DNA is longer than KLN + DNA.

The ~2 Å difference in experimental Rg_{xs} (see Table 6.2) between the KLN-DNA complexes and the KTQ-DNA complexes correlates well with the intrinsic size difference between these two polymerases. Experimentally determined Rg_{xs} values for the apo polymerases differ by ~1 Å with KLN having the larger cross-sectional radius of gyration (apoKLN – 18.5 ± 0.3; apoKTQ – 17.4 ± 0.2). Relative to the apo polymerases, the respective polymerase-DNA

complexes exhibit lower Rg_{xs} values due to the influence of the smaller cross-section of the DNA component. The measured Rg_{xs} value for each complex correlates well with the respective ACM-based Rg_{xs} value (see Table 6.2). Furthermore, length values (L in Table 6.2) calculated from the experimentally determined Rg and Rg_{xs} values correlate well with the respective Dmax values for each complex. This length calculation using Rg and Rg_{xs} is only valid for rod-like particles, and thus the excellent agreement between L and Dmax again confirms the overall rod-like shape of the polymerase-DNA complexes.

6.2.7 Comparison of Experimental and Polymerase-DNA Complex ACM-derived X-ray Scattering Parameters

Alignment of the polymerization and editing mode ACMs, as shown in Figure 6.7, reveals that the angle at which the DNA substrate protrudes from the polymerase envelope relative to the long, vertical axis is dependent upon the binding mode and differs by approximately 30°. According to the X-ray scattering parameters simulated from the ACMs and reported in Table 6.2, the binding mode does not significantly alter the Rg or Rg_{xs} but does alter the maximum particle dimension by approximately 7 Å. Since the same number of base pairs protrudes from the polymerase envelope in both the polymerization and editing mode ACM complexes, this 7 Å Dmax difference results purely from the orientational angle of the DNA substrate with respect to the polymerase. It is important to note here that if the length of the DNA protruding from the polymerase envelope in solution differs from our crystallographically based ACMs, then there will be a discrepancy between the experimental and ACM-derived Dmax values.

Comparison of the experimental and ACM-based X-ray scattering parameters reported in Table 6.2 suggests that KLN binds both pt and dsDNA in the editing mode, while KTQ binds

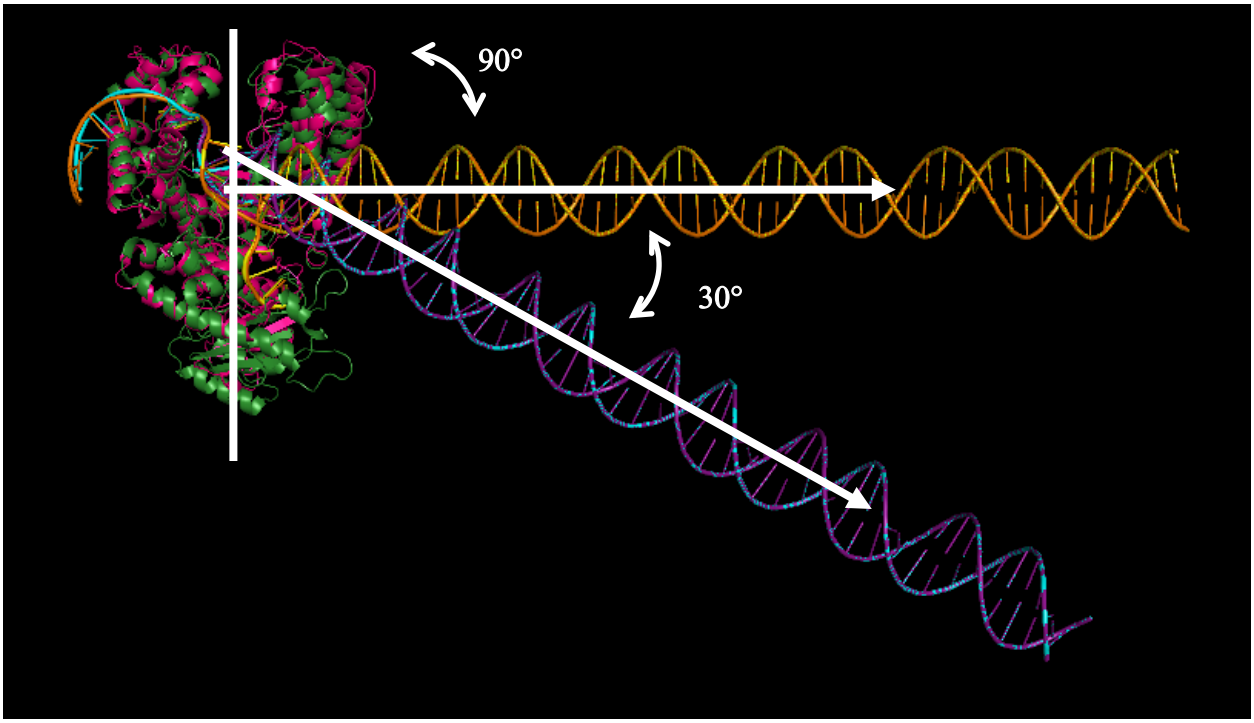


Figure 6.7. Superimposition of the ACMs from Figure 6.2 reveals significant binding mode dependent deviations in the angular relationship between the long axes of the polymerases and DNA substrates. Color designations are as denoted in Figure 6.2. ACMs were aligned according to the polymerase domains of KTQ and KLN. The vertical white line represents the long axis of the overlapped polymerases, while the white arrows correspond to the long axes of the DNA substrates bound in the editing and polymerization modes. The angles denoted are 1) between the polymerase long axis and the editing mode DNA construct and 2) between the two the DNAs complexed in the different binding modes.

both DNA structures in the polymerization mode. For the KLN + DNA complexes, the experimental Dmax and Rg are, respectively, within 6 and 2 Å of the predicted editing mode values. The KLN complexes are, however, slightly smaller than the predicted ACMs, while Klentaq complexes are much larger than expected. The larger overall dimensions of the Klentaq complexes serve to rule out the editing mode as the DNA binding mode of Klentaq under these conditions. Likewise, the difference between the polymerization mode and experimental KLN complexes is too great to be reconciled by any hypothesis based on the current crystallographic

data. Interestingly, these results also indicate that even the perfectly matched dsDNA substrate is pulled open by Klenow.

The Dmax differences between each experimentally determined complex and its respective polymerization and editing mode ACMs are graphed in Figure 6.8 along with the Dmax difference between the isolated ds-63/63mer and its ACM. This figure shows graphically that both experimental KLN-DNA complexes are closer in size (smaller Δ Dmax) to the editing mode ACM (gray bars) and that both KTQ-DNA complexes are closer in size to the polymerization mode ACMs (black bars). It is notable in Figure 6.8, however, that there are still sizable Dmax differences between 1) the experimental KTQ + DNA complexes and the respective polymerization mode ACMs and 2) the experimental DNA bound KLN complexes and the respective editing mode ACMs. These differences are at least partially explained by a farther or lesser than expected protrusion of the DNA outside the polymerase envelope, thus altering the long dimension of the complex. Two possible molecular scenarios can account for these remaining Dmax differences.

Scenario 1: In this scenario, the Dmax of each complex is purely based on the how far the DNA is pulled into the polymerase envelope. Excluding the ss template overhang of the ptDNA, all of the ACMs (Figures 6.2 and 6.7), independent of binding mode, contain ~8 base pairs within the polymerase envelope. Using a 3.4 Å rise/base pair for B-DNA, to account for the ~11-16 Å Dmax discrepancy between the experimental KTQ complexes and the respective pol mode ACMs (black bars in Figure 6.8), the DNA substrate would have to protrude ~3-5 base pairs further from the polymerase envelope, which would leave only ~3-5 base pairs interacting with KTQ within the polymerase envelope. According to this scenario, since the DNA would move further away from the active site with each additional base pair that protrudes from the

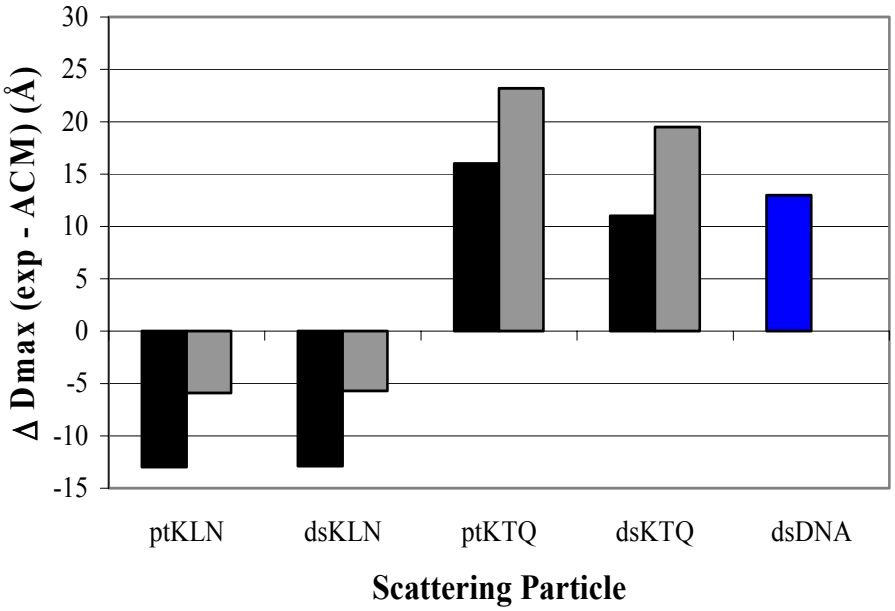


Figure 6.8. Comparison of Dmax differences between the measured polymerase-DNA complexes and their respective polymerization (black) and editing (gray) mode ACMs. The blue bar is the Dmax difference between the measured and ACM-derived values for dsDNA. The corresponding Dmax difference for ptDNA is not shown because its ACM-derived Dmax cannot be exactly predicted due to the presence of the 5' ss template overhang. Positive Δ Dmax values indicate that experimental complexes are larger than predicted from the ACMs, and vice versa for the negative Δ Dmax values. Table 6.2 Dmax values were used for all calculations.

envelope, the 3' terminus of the DNA substrate would be too far from the active site to yield a productive complex for polymerization. Thus, if this is the case, then perhaps KTQ requires additional factors, such as incoming nucleotide to induce a conformational change that would position the 3' end of DNA substrate near the active site. It is also interesting to note that to account for the even larger difference between the measured Dmax and the KTQ-edit ACMs (gray bars in Figure 6.8), an additional 7 Å or 2 bps would have to protrude from the polymerase envelope, leaving possibly only a single base pair-protein interaction. It is unlikely that a stable polymerase-DNA complex would result from a one bp interaction, thus again discounting the possibly that KTQ binds DNA in the editing mode.

Since the experimental DNA bound KLN complexes are shorter than the respective polymerization and editing mode ACMs, according to scenario 1 and in contrast to KTQ, KLN must pull the DNA further into its polymerase envelope in solution. The editing mode is well suited to account for more base pairs being pulled into the polymerase envelope because the proofreading domain could unwind more of the DNA duplex. As shown in Figures 6.2 and 6.8, the editing mode ACMs contain ~3-4 nucleotides bound within the 3' exonuclease domain. The ~6 Å Dmax discrepancy between the measured and KLN-edit ACMs corresponds to an extra ~2 bps, for a total of ~5-6 base pairs of duplex DNA pulled open by KLN and a total of ~10 bps (8 bps from ACM + 2 bps from Dmax discrepancy) pulled into the polymerase envelope. Thus, according to Scenario 1, KTQ potentially interacts with up to 7 less base pairs of DNA than KLN, and KLN melts open ~2 more base pairs than predicted by the crystallographic data.

Scenario 2: This scenario assumes that the Dmax values for isolated DNAs and polymerase-DNA complexes include ion shielding effects (discussed for the isolated DNAs in section 6.2.5). Compared to its ACM, isolated dsDNA is ~13 Å longer in Dmax and 4-6 Å larger in Rg (see Figure 6.8 and Table 6.2). The experimental Dmax values thus suggest that a 6.5 Å ion cloud equally covers each end of the DNA to yield the 13 Å longer Dmax. In the complexes, however, one end of the DNA substrate is protected by the polymerase while the other end is exposed to the solution. If a 6.5 Å ion cloud surrounds the exposed end of the DNA substrate, then the polymerization mode KTQ-DNA complexes would be ~251 Å, which is ~4-9 Å (1-3 bps) shorter than the measured Dmax's from dsKTQ and ptKTQ. Thus, for the KTQ complexes in scenario 2, a combination of counterion cloud effects and binding of ~1-3 less base pairs accounts for the discrepancy between measured and pol mode ACM-derived scattering parameters.

If similar ion shielding effects are applied to the KLN + DNA editing mode ACMs, then in solution the DNA bound KLN complexes would have a calculated Dmax of ~243 Å. The Dmax difference between the KLN-edit ACMs surrounded by a counterion layer and the experimental KLN complexes is then ~13 Å, which corresponds to ~4 bp of B-DNA. Thus, this scenario suggests that KLN denatures ~7-8 base pairs of the DNA duplex because 3-4 nucleotides are already predicted to be pulled open in the KLN-edit ACMs. In scenario 2, the total number of bps not protruding from the polymerase envelope along the long axis of the DNA substrate is ~12 bases (i.e. 8 bps predicted + 4 from Dmax difference). However, it is unlikely that all 12 base pairs will remain within the polymerase envelope; the proofreading domain (due to its size) can probably only contain ~5-6 nucleotides, and the remaining 2-3 denatured nucleotides would most likely exit the proofreading domain (which would not contribute to the long dimension of the complex). Thus, according to scenario 2, KLN likely protects 2-5 more base pairs than KTQ, and melts open up to 4 more base pairs than predicted by crystallographic data.

While residual exonucleolytic cleavage of some of these nucleotides by Klenow's disabled exonuclease site cannot be completely ruled out, comparative analysis of the forward scattering intensities of the complexes (I(0)) does not support the presence of such activity (i.e. the I(0)'s and their resulting molecular weights scale as expected according to the size and concentration of the complexes – ptKLN > dsKLN > ptKTQ > dsKTQ; the differences between the I(0) (MW) values of ptKLN and dsKLN like those of ptKTQ and dsKTQ can be accounted for by the additional 7 ss nucleotides of the ptDNA 5' ss template overhang).

In either scenario 1 or 2, KLN binds more base pairs than KTQ and this result correlates well with salt dependence studies that demonstrate a smaller net ion release upon binding DNA

for KTQ than KLN and which also suggest that KTQ interacts with less DNA base pairs [81]. It bears reemphasis that according to either of the proposed scenarios, editing mode binding by KTQ and polymerization mode binding by KLN cannot be reconciled.

6.2.8 Shape Modeling of the Polymerase-DNA Complexes and Isolated DNA Constructs

Examination of the various scattering curve transformations and quantitative analysis of the X-ray scattering parameters indicates that the global conformations of polymerase-DNA complexes are comprised of both globular and rod-like structural features, and that the isolated DNA constructs are similar to rod-like B-DNA. Solution structures of protein-DNA complexes such as these, in which the DNA component is very long and protrudes far outside of the protein envelope, have rarely been experimentally examined. Thus, *ab initio* shape models of the polymerase-DNA complexes and the isolated DNAs were generated from the experimental scattering data.

Only the SAXS curves were used to assess concentration dependence and determine the Rg, I(0), and Rg_{xs} X-ray scattering parameters. For *ab initio* shape modeling, however, the lowest concentration SAXS curve was merged with the highest concentration WAXS curve (as described in Chapter 3). Combining the scattering curves in this way yields a final scattering curve for modeling that consists of both 1) high quality data at wider angles, and 2) small angle data that is relatively free from concentration dependent behavior. This approach was deemed more accurate for our data sets than attempting to extrapolate a theoretical scattering curve at infinite dilution. As expected, GNOM analysis of the merged scattering curves, in preparation for *ab initio* shape modeling, yielded Rg and I(0)/c values that were comparable to those obtained for the respective unmerged lowest concentration SAXS curves.

The GNOM-analyzed merged X-ray scattering curves were each truncated at a scattering vector of approximately 0.3 $Å^{-1}$ and then subjected to *ab initio* bead modeling, as described in Chapter 3, using the computer program DAMMIF [107]. Ten to twelve individual models were created from each scattering curve to account for degeneracy in the bead modeling process. In other words, a number of similar models can fit the scattering data. Examination of and averaging of 10-12 individual models has been deemed a representative sample of the entire ensemble of degenerate solutions/models. For comparison to the experimental studies, similar modeling exercises were performed using CRYSOL-simulated X-ray scattering curves from the dsDNA and dsKTQ-pol ACMs.

As demonstrated in Figure 6.9, all of the experimental X-ray scattering curves are extremely well fit by the bead models. Standard deviation on the average χ value for goodness of fit is ≤ 0.03 for each experimental X-ray scattering curve (see Table 6.3). Six out of ten individual ptKTQ bead models are shown in Figure 6.10 along with the averaged and filtered molecular envelopes for this data set. Both globular and extended rod-like domains, corresponding to the polymerase and the DNA components of the complex, are readily apparent features of each individual bead model. These bead model characteristics are extremely stable and are still clearly resolved in the averaged and filtered molecular envelope. However, the individual features of each component, such as the curvature of the rod-like domain or the hole in the globular domain, are not very consistent from one individual bead model to the next. The curvature and bulges within the rod-like domain, which makes up more than half of the total length of the complex, dominate the alignment of the individual bead models. Thus, some features of the globular domain are lost during averaging. As for the rod-like portion of the

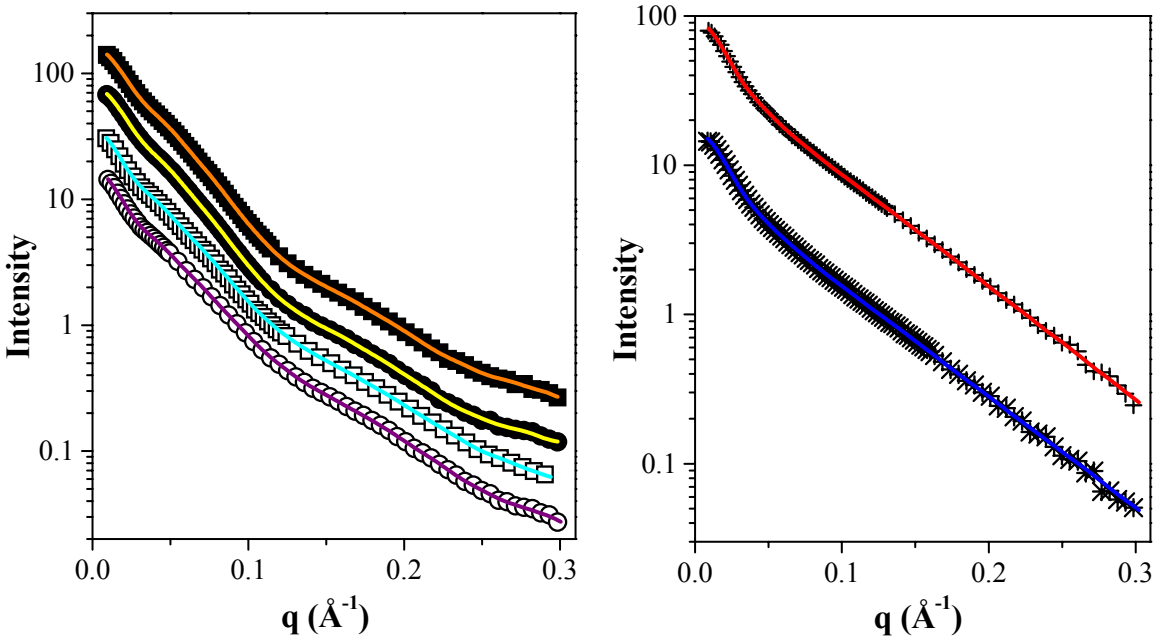


Figure 6.9. Representative fits of individual DAMMIF bead model-derived scattering curves to experimental X-ray scattering patterns. Experimental X-ray scattering curves for ptKLN (■), dsKLN (●), ptKTQ (□), and dsKTQ (○), ptDNA (+), and dsDNA (✳) are superimposed with the corresponding best fit (i.e. best χ) bead model-derived scattering pattern (solid lines). Color designations are the same as those used in Figure 6.6. For clarity, measured scattering curves and fitted data are arbitrarily shifted along the y-axis, and only every second or third data point is shown.

Table 6.3. Parameters describing quality of DAMMIF bead models.

Isolated DNA or Complex[a]	Mean Goodness of fit (χ)[b]	Mean Normalized Spatial Discrepancy (NSD)[c]
ptKTQ	1.39 ± 0.01	1.04 ± 0.06
dsKTQ	1.22 ± 0.02	1.12 ± 0.10
ptKLN	1.52 ± 0.02	1.04 ± 0.02
dsKLN	1.43 ± 0.01	0.95 ± 0.05
ptDNA	0.98 ± 0.01	0.81 ± 0.03
dsDNA	1.12 ± 0.03	1.00 ± 0.03

[a] 12 individual models were created for each isolated DNA, and 10 individual models for each complex.
[b] Describes fit between bead model-derived scattering and experimental scattering; mean ± standard deviation for each set of individual bead model.
[c] Describes quality of the match between the individual bead models of each set.

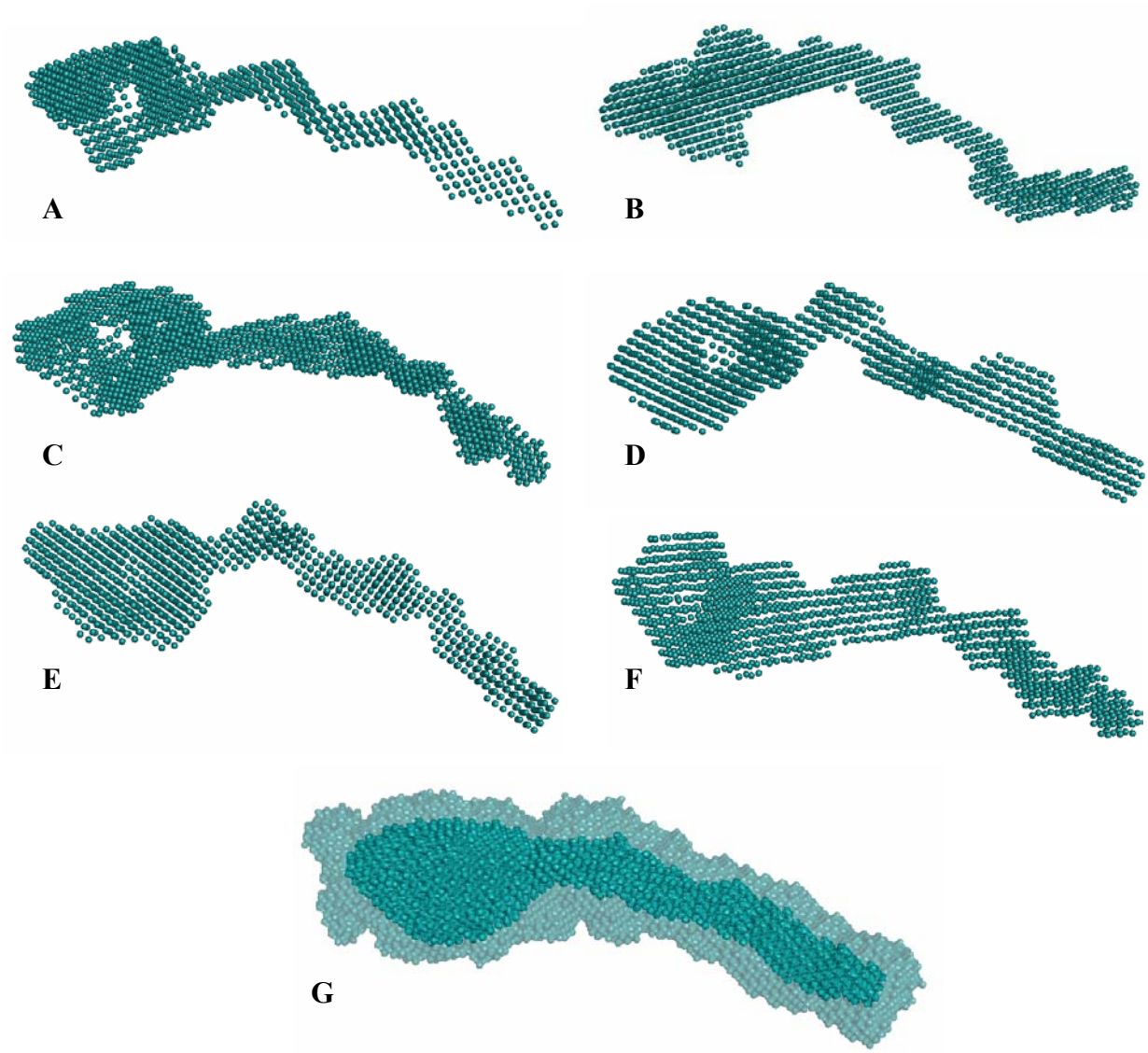


Figure 6.10. Representative experimentally determined DAMMIF bead models. Individual bead models A – F were each calculated via a separate DAMMIF run using the same measured X-ray scattering data for ptKTQ (only 6 out of 10 individual models are shown in this figure, but the other 4 bead models exhibit similar structural details). Panel G contains the larger averaged bead model superimposed with the final, filtered molecular envelope, which is the averaged model filtered to the mean excluded volume of the scattering particle determined from the individual bead models. The averaged bead model (larger perimeter of model G) represents the total volume occupied by all of the individual, aligned bead models. Major structural features are retained by the final filtered shape model (smaller imbedded perimeter in model G), while inconsistent characteristics of the discrete bead models are eliminated.

models, the averaging and filtering procedure typically smoothes out the small bulges, but retains any consistent global curvature (see panel G of Figure 6.10).

The final averaged, filtered molecular envelopes for each polymerase-DNA complex are shown in Figure 6.11 along with the dsKTQ-pol and dsKLN-edit ACMs displayed in a surface representation scaled similar to that of the experimentally determined molecular envelopes. The overall shape similarity between the empirical molecular envelopes and the ACMs is quite striking, and the experimental shape models confirm that the polymerase forms a stable complex with the DNA substrate by binding at a DNA terminus as opposed to binding somewhere in the middle of the long DNA construct. Although this characteristic of the polymerase-DNA complexes is also indicated by the scattering curve transformations and assessment of quantitative scattering parameters (Table 6.2), it is much more apparent in the shape models. The same polymerase-specific Rg and Dmax differences between the KTQ + DNA and KLN + DNA complexes as noted in section 6.2.6 are also observed in the these shape models; i.e. according to the shape models, the DNA bound KTQ complexes are ~30 Å longer in Dmax, and ~8 Å larger in Rg.

A visually obvious difference in the solution shape models of the four polymerase-DNA complexes is the variation in the degree of curvature within the rod-like domain. The dsKTQ and ptKLN molecular envelopes in Figure 6.11 exhibit the least amount of curvature, while the ptKTQ and dsKLN shape models display more significant apparent bending of their DNA substrates. To determine if this global curvature of the DNA is reflective of actual bending of the DNA in solution, we created shape models of the pt and dsDNA constructs from the measured X-ray scattering curves and compared the resulting structures with a shape model of the simulated dsDNA ACM (which is perfect B-DNA in a rigid rod conformation) generated

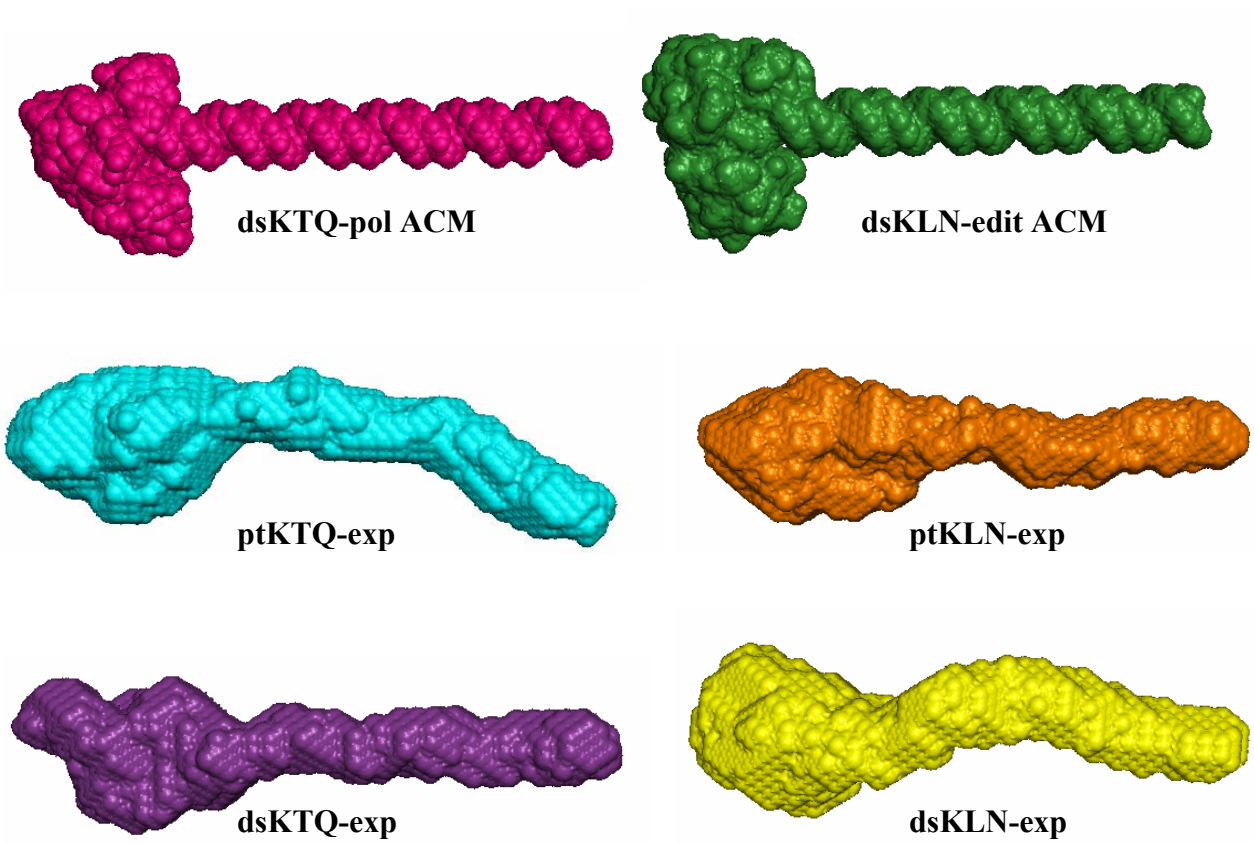


Figure 6.11. The experimentally determined polymerase-DNA complex molecular envelopes share a significant likeness in general overall appearance to the global shape of the polymerase-DNA ACMs. The dsKTQ-pol and dsKLN-edit ACMs *(top)* and the final averaged, filtered shape models *(middle and bottom)* of ptKTQ, ptKLN, dsKTQ, and dsKLN are each displayed in similarly scaled surface representation with the van der Waals sphere size set to 5 Å. The X-ray scattering-derived shape envelopes are *ab initio* models, and utilize no crystallographic or other information outside of the scattering data and a cylindrical search volume.

from the theoretical X-ray scattering curve. The filtered, averaged shape models of the measured ptDNA and dsDNA constructs and the simulated dsDNA ACM are shown in Figure 6.12 aligned with the dsDNA atomic coordinate model. The ptDNA shape envelope superimposed with the ptDNA atomic coordinate model is not shown because this alignment is skewed by the presence of the 5' ss template overhang, for which no corresponding volume can be identified in the shape model.

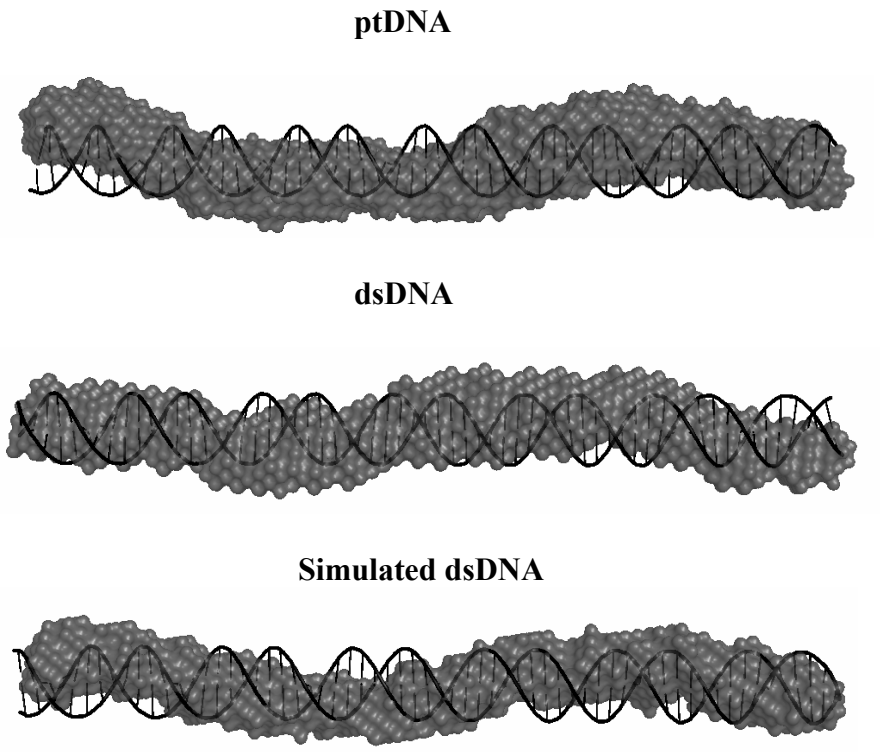


Figure 6.12. Averaged, filtered DAMMIF shape models of isolated DNA constructs. Shape models (gray, surface representation) created from pt and dsDNA experimental X-ray scattering curves and simulated dsDNA scattering data are each superimposed with the simulated dsDNA atomic coordinate model (black).

The scattering curves from the individual bead models of the experimental DNA constructs yielded excellent fits to the measured X-ray scattering data (see Figure 6.9 and Table 6.3). Also, the individual bead models each exhibit a similar global conformation to the other bead models of the set, as evidenced by the low NSD values for the pt and dsDNA constructs (see Table 6.3). Since the mean NSD values for the isolated DNAs are lower than the mean NSD for the set of ptKTQ individual bead models, the isolated DNA individual bead models of each set are at least as similar as or more similar to one another than the representative individual bead models of ptKTQ shown in Figure 6.10. Like the ptKTQ individual bead models, the isolated DNA individual models display many random bulges and bends that are smoothed out during averaging and filtering process. However, as illustrated in Figure 6.12 some global

curvature is retained in the averaged, filtered shape envelopes. Aside from disagreement with the curvature, the dsDNA ACM superimposes well with the experimental averaged, filtered molecular envelopes of both the pt and dsDNAs, again indicating that these two DNA constructs have very similar solution conformations and maximum particle dimensions.

Although slight bending of the DNA molecules in solution would not be unexpected, the curvature observed in the experimental shape models of the isolated DNAs and complexes cannot be considered to reflect actual curvature of these DNA constructs in solution because, as shown in Figure 6.12 (bottom panel), even the shape model created from simulated B-DNA with a strict rigid rod conformation also displays curvature. In fact, when the averaged, filtered molecular envelopes are compared, the curvature in the experimentally determined DNA shape models is nearly identical to that of the ACM-based envelope. Also, Svergun and colleagues assert that simulated rigid rod-like particles cannot be reconstructed without curvature by this bead modeling algorithm [134].

It bears mentioning, however, that the individual bead models from simulated data are typically straighter than individual models from experimental data (for both complex and the isolated DNA comparisons against simulated data from the respective ACMs). Also, a periodicity that seems to often cause the bulges is observed in the individual models, but is smoothed over by the averaging and filtering procedure. This periodicity is well correlated with the distance per helical turn of B-DNA, which is ~34 Å. Further studies of the individual bead models will be necessary to discern whether these characteristics of the individual bead models (which are more detailed than the final averaged, filtered shape model) truly reflect properties of the DNA in solution or are purely artifacts of the bead modeling algorithm.

Another notable feature of the experimentally determined *ab initio* shape envelopes of the pt and dsDNA constructs is that they are slightly fatter (i.e. wider cross-section) than the shape model from the simulated dsDNA scattering. This result correlates well with the higher D_{xs} observed in the comparative quantitative analysis of the scattering parameters of the isolated DNAs in solution versus the simulated dsDNA ACM. Conversely, the corresponding Dmax difference between experimental and simulated DNAs is not clearly visible in the shape envelopes due to a combination of the curvature within the DNA shape models and the difficulty of accurately representing these 3-dimensional shape envelopes in 2-dimensions. However, when the three shape envelopes and their overlapped ACMs are viewed in 3D, regions of excess volume near the DNA termini are observed for the experimental shape models but not for the shape envelope from the simulated data. These regions of excess volume add ~6 Å to each end of the dsDNA ACM, and were earlier discussed as being likely due to counterion association.

6.2.9 Superimpositioning of the Polymerase-DNA Complex Shape Models with the ACMs

6.2.9.1 Determination of the "Best Possible" Alignment Using Simulated X-ray Scattering Data

According to the quantitative, comparative analysis of the experimental X-ray scattering parameters, it is clear that our crystallographically-based ACMs do not completely describe the solution structures of the polymerase-DNA complexes as seen by X-ray scattering. Thus, prior to aligning our shape models and ACMs, we determined the expected best possible alignment by superimposing the dsKTQ-pol ACM with the averaged, filtered shape envelope obtained from the simulated scattering data generated from this same ACM. As observed for all modeling runs, using experimental or simulated data, the individual bead models from the dsKTQ-pol simulated scattering exhibited random bending and curvature within the rod-like domain. However, as expected for models generated from simulated data, the DNA component was on average much

straighter than observed in the individual models from experimental data. In fact, out of all of the modeling runs performed (including those performed with simulated scattering data from the isolated dsDNA ACM), the individual bead models from simulated dsKTQ-pol exhibited the most stable/less degenerate set of structures.

The alignment between the shape model generated from the simulated data and the dsKTQ-pol mode ACM is shown in Figure 6.13. Although the superimposition is very good (NSD = 1.073), it is not perfect even though the molecular envelope was generated from "perfect" simulated scattering data. Problem areas in the averaged, filtered shape envelope include 1) subtle curvature within the rod-like domain and 2) insufficient volume within the globular domain to account for the tips of the fingers, thumb, and proofreading domains. Problem area #1 was expected according to the modeling results of simulated dsDNA ACM (Figure 6.12). However, problem area #2 was unexpected.

Two primary causes for the lack of volume within the globular domain can be identified: 1) as previously mentioned, the rod-like domain of the individual bead models dominates the alignment of the individual models during the averaging process. Thus, sometimes even stable features of the globular domain are lost at the expense of aligning the rod-like domain. Introduction of a weighting factor into the alignment algorithm, which might allow the user to increase or decrease the importance of certain regions of the bead model known to be error prone, could decrease the loss of information within the globular domain. While this is certainly a possible cause for some of the missing volume in the final shape envelope, it cannot be the only source of the problem because even the individual bead models do not have enough volume within the globular domain to sufficiently account for the entire polymerase.

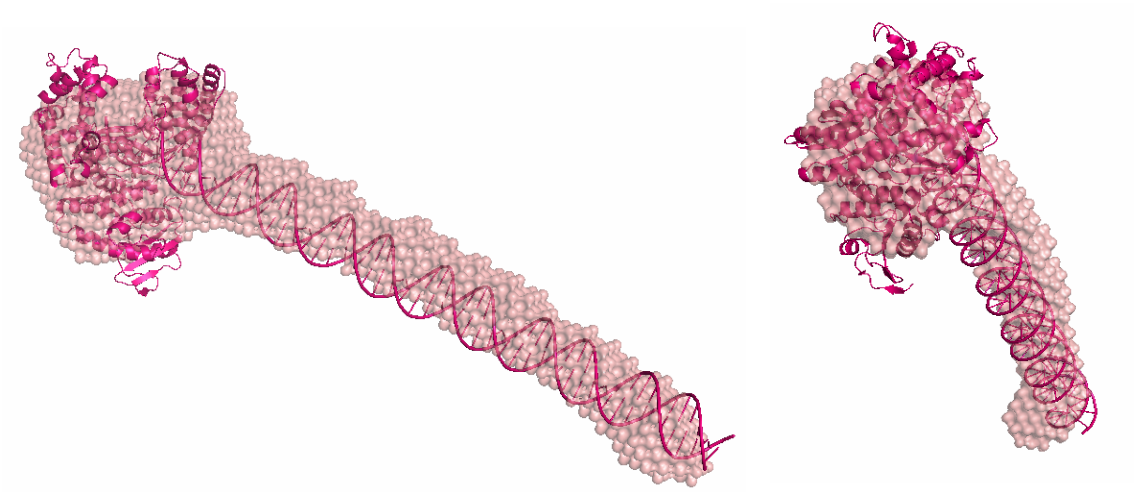


Figure 6.13. DAMMIF shape model generated from the CRYSOL-simulated X-ray scattering curve for the dsKTQ-pol ACM. The averaged and filtered DAMMIF molecular envelope (pink, transparent surface representation) is superimposed with the ptKTQ-pol ACM (hot pink, cartoon representation). *Right:* Alternate view, model and ACM are rotated 90° with respect to the y-axis.

2) The DAMMIN and DAMMIF algorithms [107] assume a homogeneous electron density of the scattering particle, which is certainly not the case for the polymerase-DNA complexes because the electron density of the DNA is higher than that of the protein (see Table 3.4). For protein-DNA complexes in which the DNA is small relative to the protein and mostly protected within the protein envelope, this difference in electron density contrast is generally considered to be negligible. However, for our complexes, the protein and DNA components are nearly equivalent in size, and thus this factor cannot be completely ignored. Since the electron density of DNA is higher than that of protein and since the total length and width of the DNA component within the shape model shown in Figure 6.13 are well accounted for but the protein region is deficient in volume, the inability to account for the differential electron density contrast within the macromolecular complex is likely the greatest source of this lack of protein volume. Recently a new program, MONSA [161], was developed that targets this exact problem – *ab*

initio modeling of complexes composed of macromolecules with differential electron density contrasts, but since this more advanced modeling process is both non-trivial and significantly more computationally intensive, generating *ab initio* models using this program will have to be the focus of future research.

Another simulated data test/benchmark involved superimposing the simulated dsKTQ-pol shape envelope with the dsKTQ-edit mode ACM to determine whether or not the resolution of the shape model is sufficient to distinguish between the polymerization and editing mode ACMs for the same polymerase. The averaged, filtered dsKTQ-pol shape envelope superimposes equally well, possibly slightly better, with the dsKTQ-edit ACM (NSD = 1.063) as with the dsKTQ-pol ACM. Therefore, although the quantitative, comparative analysis of the X-ray scattering parameters can distinguish between the polymerization and editing modes, this simulation shows that the shape envelope topology alone cannot resolve between these two binding modes.

6.2.9.2 Alignment of the Polymerase-DNA Complex ACMs with the Experimentally Determined Molecular Envelopes

Although the simulation results indicate that the binding modes could not be experimentally distinguished solely based on shape model alignments, we examined the superimposition of each experimentally determined shape envelope against its respective polymerization and editing mode ACM and confirmed that neither topological alignment is significantly better or worse than the other. NSD differences between the two binding modes were less than 0.09 (all NSDs for each experimentally determined shape envelope aligned with each binding mode were between 1.15 and 1.4). Thus, each experimental molecular envelope is shown in Figure 6.14 superimposed with either its respective polymerization or editing mode ACM based on the conclusions from the quantitative, comparative analysis of the experimental

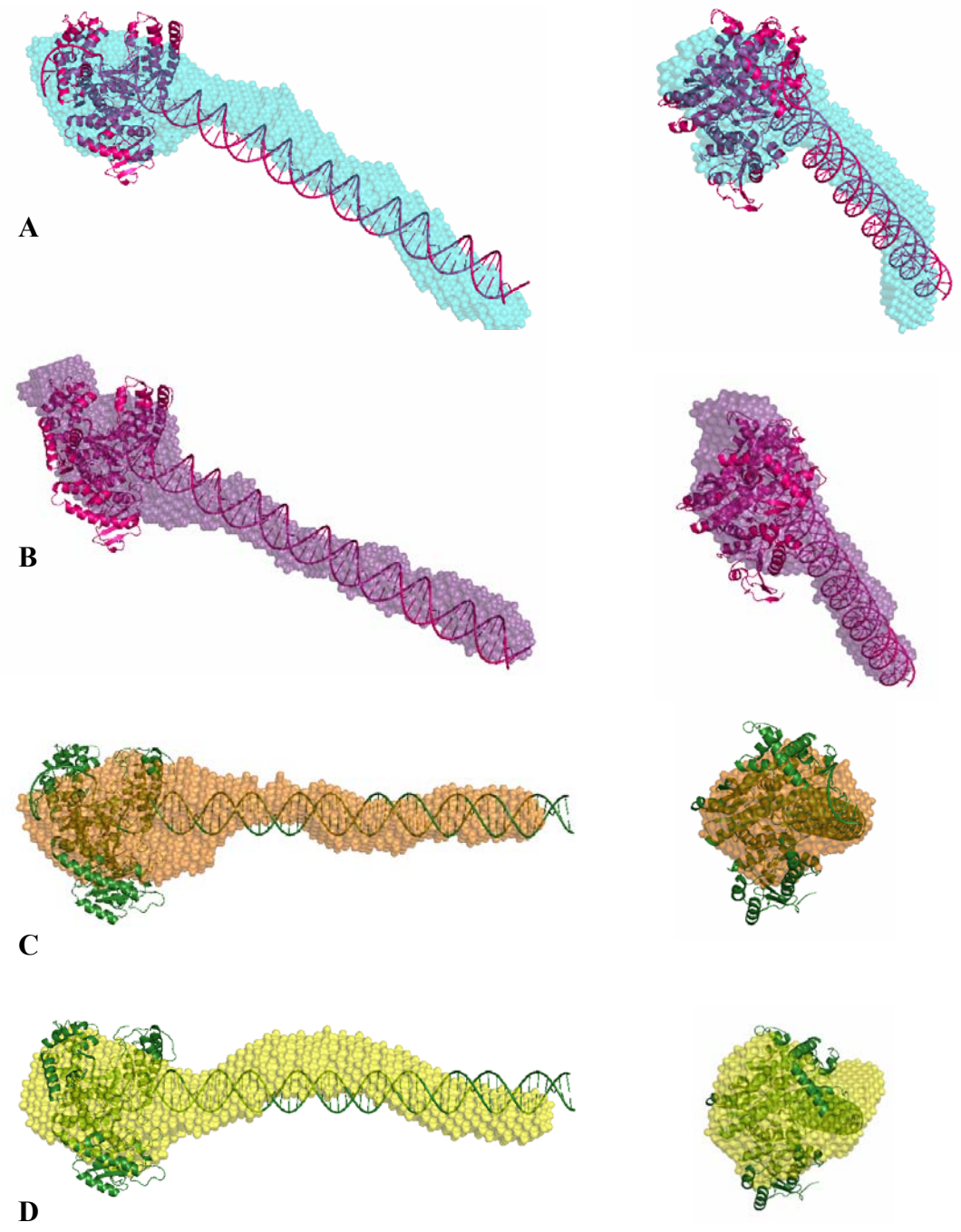


Figure 6.14. Proposed DNA binding mode for each polymerase-DNA complex based on comparison of experimental and ACM-derived X-ray scattering parameters. The averaged, filtered experimentally determined molecular envelopes (colored, transparent surface representations) for ptKTQ *(A)*, dsKTQ *(B)*, ptKLN *(C)*, and dsKLN *(D)* are respectively superimposed against ptKTQ-pol, dsKTQ-pol, ptKLN-edit, and dsKLN-edit ACMs (cartoon representations). Right: Alternate view, rotated 90° about the vertical axis.

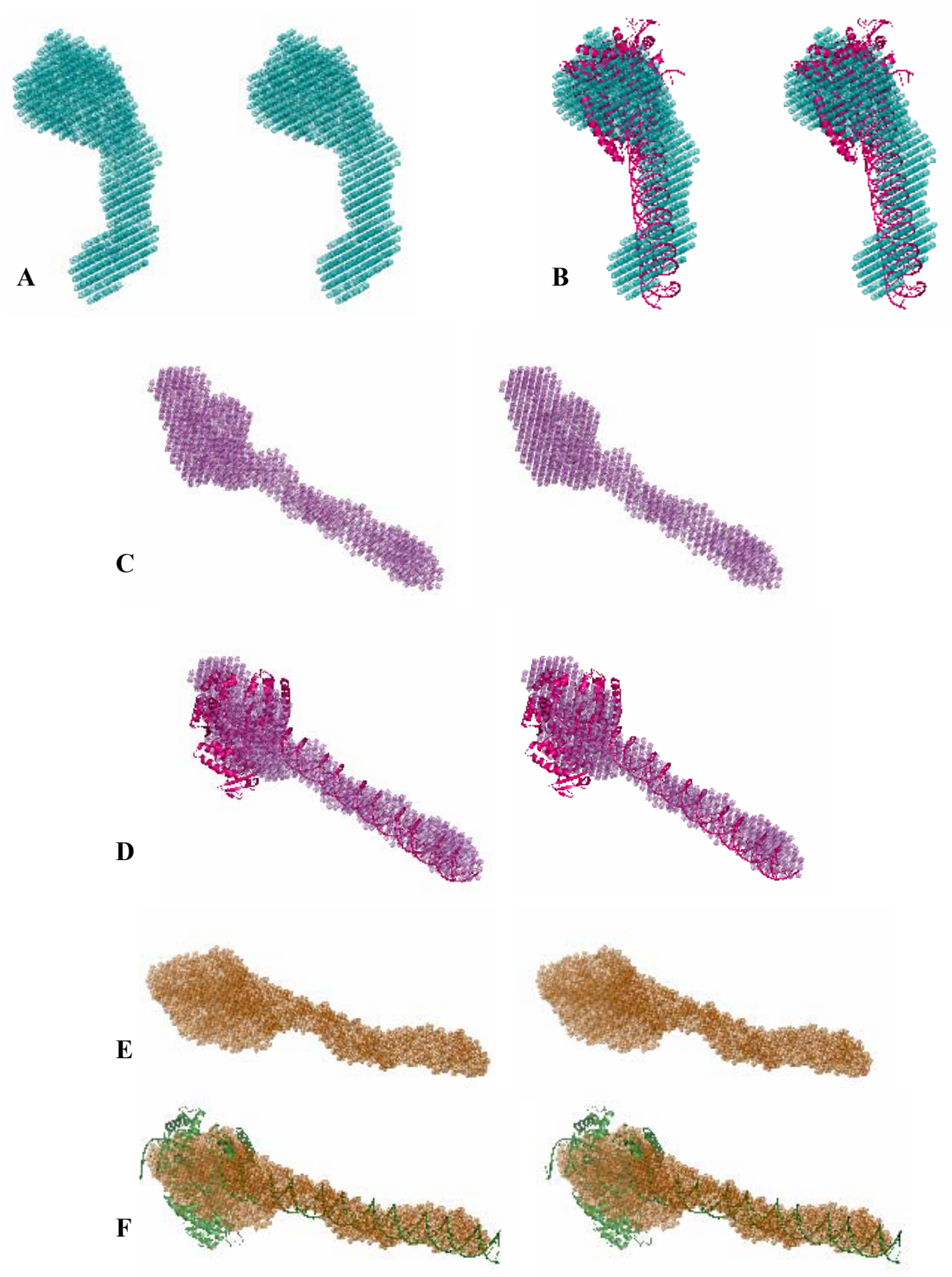


Figure 6.15. Stereo diagrams of the experimentally determined polymerase-DNA complex shape models. Shape models for ptKTQ (A and B), dsKTQ (C and D), and ptKLN (E and F) are shown both with and without a corresponding docked ACM (ptKTQ-edit, dsKTQ-pol, and ptKLN-edit).

and ACM-derived X-ray scattering parameters (i.e. ptKTQ and dsKTQ are superimposed with the respective polymerization mode ACM, while ptKLN and dsKLN are superimposed with the respective editing mode ACM). Since these 3-dimensional envelopes are difficult to adequately represent in 2 dimensions, Figure 6.15 also shows the shape envelopes of ptKTQ, dsKTQ, and ptKLN, here represented in stereo and in various orientations; each shape model in Figure 6.15 is shown both with and without an overlapped ACM.

As predicted by the simulated modeling exercise in Figure 6.13, superimposition of the experimental shape models and ACMs reveals that each polymerase component cannot be completely accounted for by the experimentally determined shape envelope, and that the missing volume is localized to the edges of the long dimension of the polymerase (see Figure 6.14). Two observations from the experimental polymerase-DNA complex shape models and ACM-alignments that require further investigation include: 1) The dsKTQ shape envelope is the suspicious. While the general overlap against the dsKTQ-pol ACM is relatively good, the bulge above the fingers subdomain cannot be easily explained. Although the set of individual dsKTQ bead models exhibited the poorest stability (as evidenced by the highest NSD in Table 6.3), this bulge was quite consistent among the individual bead models. 2) The dsKLN individual shape models have an unexpectedly large volume that is ~1.3 times higher than the other that of the other DNA bound polymerase complexes. It is also apparent from visual comparison of the shape models in either Figure 6.11 or 6.14 that the rod-like domain of the dsKLN shape envelope is slightly fatter than for the other polymerase-DNA complex shape models. Since the global conformation of dsKLN in solution exhibits a molecular envelope that is relatively consistent with the other polymerase-DNA complexes in its overall features, it is difficult to conclude from the results of this X-ray scattering investigation what might be the cause of this increased

volume. It is possible that because the lowest concentration SAXS curve was used for the modeling that this curve for dsKLN is affected by aggregation/oligomerization as indicated by the concentration dependence of dsKLN in Figure 6.4. However, potential residual structure factor information within the lowest concentration X-ray scattering curves used for *ab initio* shape modeling does not appear to negatively affect the molecular envelopes of ptKTQ, dsKTQ, and ptKLN.

Nevertheless, analysis of the ACM-superimposed shape envelopes yields results that are consistent with the comparative, quantitative assessment of the X-ray scattering parameters. 1) In solution, DNA bound KTQ is larger than its polymerization mode ACM, and larger than DNA bound KLN. The ptKTQ shape model is ~9 Å longer at the distal end of the DNA substrate, while its overall Dmax is ~260 Å. 2) The KLN-DNA shape envelopes are shorter than their respective ACMs, and again suggest that KLN pulls an additional ~4 base pairs into its polymerase envelope, likely by binding the DNA substrate in the editing mode and melting the DNA duplex. These results from the shape model-ACM alignments seem to be more consistent with Scenario 2 of the comparative, quantitative assessment.

6.3 Concluding Summary

The molecular envelopes of the polymerase-DNA complexes and the isolated DNAs presented in this chapter are the among the first solution scattering-based *ab initio* shape models of long rod-like DNAs, either in isolation or protein-bound. Several interesting observations have resulted from this X-ray scattering characterization of the global conformations of Klentaq and Klenow bound to the two different DNA structures, pt and dsDNA. All of the polymerase-DNA complexes exhibit 1:1 stoichiometry in solution at least at low concentration. The dsKLN and ptKLN complexes, however, display dissimilar Rg and I(0)/c concentration dependencies,

which may be a consequence of recently identified thermodynamic differences between Klenow binding to ptDNA versus dsDNA (Wowor and LiCata, unpublished). In each complex, the polymerase binds near the terminus of the DNA duplex, and the portion of the DNA duplex that extends out of the polymerase envelope into solution maintains a relatively rigid rod-like structure. While the global conformations of DNA bound Klentaq and Klenow exhibit a clear dependence on the polymerase component, the overall solution structures are not significantly altered by the structure of the DNA substrate, at least at low concentrations. The quantitative, comparative assessments of the experimental and ACM-derived scattering parameters indicate that KTQ-DNA forms a complex that is ~30 Å longer in Dmax and ~8 Å larger in Rg than the KLN-DNA complexes, and that KTQ binds both DNA substrates in the polymerization mode, while KLN binds both substrate structures in the editing mode. Both the averaged, filtered molecular envelopes and the comparative, quantitative analyses reveal that the exact details of the DNA binding topology of the polymerases may differ from the crystallographically-derived ACMs. In solution DNA bound KLN is shorter than its respective editing mode ACM, while DNA bound KTQ is longer than its respective polymerization mode ACM. Potential explanations for these results are discussed in the context of two scenarios based on topological binding differences in solution which persist both in the absence and presence of a presumed SAXS detectable counterion cloud surrounding the negatively charged DNA.

A detailed characterization of X-ray scattering-derived *ab initio* shape models of protein-DNA complexes such as these is a difficult task with uncertainties arising from the inhomogeneous electron density contrast of the protein-DNA complex and the counterion environment of the DNA. A combination of X-ray scattering and neutron scattering is considered most powerful for studying the solution structure of protein-DNA complexes.

However, some proteins are adversely affected by the D_2O used for contrast matching in neutron scattering. While our results indicate that valuable structural information can be obtained for protein-DNA complexes from X-ray scattering alone, extending these SAXS/WAXS measurements with neutron scattering should be a future aim of these studies. Also, with more information concerning the counterion cloud around the DNA and an ability to account for the differential electron density contrast of the protein and DNA components, a more detailed picture of the protein-DNA complex may be obtained from the X-ray scattering data alone.

Ab initio bead modeling of X-ray scattering data is still a relatively new methodology, and is sometimes viewed with skepticism due to the fact that multiple degenerate shape models will fit the experimental scattering curve equally well. Each study, such as this one, demonstrating remarkable similarity between the *ab initio* scattering-derived shape models and the high resolution crystal structures validates the use of this technique to determine the global conformation of macromolecules, which exhibit altered solution states or for which no high resolution structures exist.

REFERENCES

1. Bessman, M.J., Kornberg, A., Lehman, I.R. and Simms, E.S. (1956) Enzymic synthesis of deoxyribonucleic acid. *Biochim. Biophys. Acta* **21,** 197-198.

2. Lehman, I.R. (2003) Discovery of DNA polymerase. *J. Biol. Chem.* **278,** 34733-34738.

3. Lehman, I.R., Zimmerman, S.B., Adler, J., Bessman, M.J., Simms, E.S. and Kornberg, A. (1958) Enzymatic synthesis of deoxyribonucleic acid. V. Chemical composition of enzymatically synthesized deoxyribonucleic acid. *Proc. Natl. Acad. Sci. USA* **44,** 1191-1196.

4. Kornberg, A. (1969) Active center of DNA polymerase. *Science* **163,** 1410-1418.

5. Kornberg, A. (1988) DNA replication. *J. Biol. Chem.* **263,** 1-4.

6. Kornberg, A. and Baker, T. (1992) *DNA replication*, 2nd ed., Freeman: New York.

7. Garrett, R.H. and Grisham, C.M. (2005) *Biochemistry*, 3rd ed., Thomson Brooks/Cole: Belmont, 898-941.

8. Rothwell, P.J. and Waksman, G. (2005) Structure and mechanism of DNA polymerases. *Adv. Protein. Chem.* **71,** 401-440.

9. Braithwaite, D.K. and Ito, J. (1993) Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res.* **21,** 787-802.

10. Delarue, M., Poch, O., Tordo, N., Moras, D. and Argos, P. (1990) An attempt to unify the structure of polymerases. *Protein Eng.* **3,** 461-467.

11. Ito, J. and Braithwaite, D.K. (1991) Compilation and alignment of DNA polymerase sequences. *Nucleic Acids Res.* **19,** 4045-4057.

12. Joyce, C.M. and Steitz, T.A. (1994) Function and structure relationships in DNA polymerases. *Annu. Rev. Biochem.* **63,** 777-822.

13. Joyce, C.M. and Steitz, T.A. (1995) Polymerase structures and function: variations on a theme? *J. Bacteriol.* **177,** 6321-6329.

14. Steitz, T.A. (1999) DNA polymerases: structural diversity and common mechanisms. *J. Biol. Chem.* **274,** 17395-17398.

15. Lehman, I.R. and Uyemura, D.G. (1976) DNA polymerase I: essential replication enzyme. *Science* **193,** 963-969.

16. Perler, F.B., Kumar, S. and Kong, H. (1996) Thermostable DNA polymerases. *Adv. Protein Chem.* **48,** 377-435.

17. Steitz, T.A. (1998) A mechanism for all polymerases. *Nature* **391,** 231-232.

18. Saturno, J., Lazaro, J.M., Esteban, F.J., Blanco, L. and Salas, M. (1997) ϕ29 DNA polymerase residue Lys383, invariant at motif B of DNA-dependent polymerases, is involved in dNTP binding. *J. Mol. Biol.* **269,** 313-325.

19. Johnson, A. and O'Donnell, M. (2005) Cellular DNA replicases: components and dynamics at the replication fork. *Annu. Rev. Biochem.* **74,** 283-315.

20. Lamers, M.H., Georgescu, R.E., Lee, S.G., O'Donnell, M. and Kuriyan, J. (2006) Crystal structure of the catalytic alpha subunit of *E. coli* replicative DNA polymerase III. *Cell* **126,** 881-892.

21. Bailey, S., Wing, R.A. and Steitz, T.A. (2006) The structure of *T. aquaticus* DNA polymerase III is distinct from eukaryotic replicative DNA polymerases. *Cell* **126,** 893-904.

22. Chien, A., Edgar, D.B. and Trela, J.M. (1976) Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *J. Bacteriol.* **127,** 1550-1557.

23. Lawyer, F.C., Stoffel, S., Saiki, R.K., Myambo, K., Drummond, R. and Gelfand, D.H. (1989) Isolation, characterization, and expression in *Escherichia coli* of the DNA polymerase gene from *Thermus aquaticus*. *J. Biol. Chem.* **264,** 6427-6437.

24. Brock, T.D. (1974), in *Bergey's Manual of Determinative Bacteriology*, edited by R.E. Buchanan and N.E. Gibbons, pp. 285, Williams and Wilkins: Baltimore.

25. Patel, P.H. and Loeb, L.A. (2000) DNA polymerase active site is highly mutable: evolutionary consequences. *Proc. Natl. Acad. Sci. USA* **97,** 5095-5100.

26. Karantzeni, I., Ruiz, C., Liu, C.C. and Licata, V.J. (2003) Comparative thermal denaturation of *Thermus aquaticus* and *Escherichia coli* type 1 DNA polymerases. *Biochem. J.* **374,** 785-792.

27. Mullis, K.B. and Faloona, F.A. (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.* **155,** 335-350.

28. Joyce, C.M. and Grindley, N.D. (1983) Construction of a plasmid that overproduces the large proteolytic fragment (Klenow fragment) of DNA polymerase I of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **80,** 1830-1834.

29. Klenow, H. and Henningsen, I. (1970) Selective elimination of the exonuclease activity of the deoxyribonucleic acid polymerase from *Escherichia coli* B by limited proteolysis. *Proc. Natl. Acad. Sci. USA* **65,** 168-175.

30. Kim, Y., Eom, S.H., Wang, J., Lee, D.S., Suh, S.W. and Steitz, T.A. (1995) Crystal structure of *Thermus aquaticus* DNA polymerase. *Nature* **376,** 612-616.

31. Alba, M. (2001) Replicative DNA polymerases. *Genome Biol.* **2,** reviews3002.

32. Uyemura, D., Eichler, D.C. and Lehman, I.R. (1976) Biochemical characterization of mutant forms of DNA polymerase I from *Escherichia coli*. II. The polAex1 mutation. *J. Biol. Chem.* **251,** 4085-4089.

33. Korolev, S., Nayal, M., Barnes, W.M., Di Cera, E. and Waksman, G. (1995) Crystal structure of the large fragment of *Thermus aquaticus* DNA polymerase I at 2.5-A resolution: structural basis for thermostability. *Proc. Natl. Acad. Sci. USA* **92,** 9264-9268.

34. Beese, L.S., Friedman, J.M. and Steitz, T.A. (1993) Crystal structures of the Klenow fragment of DNA polymerase I complexed with deoxynucleoside triphosphate and pyrophosphate. *Biochemistry* **32,** 14095-14101.

35. Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G. and Steitz, T.A. (1985) Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature* **313,** 762-766.

36. Brautigam, C.A. and Steitz, T.A. (1998) Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.* **8,** 54-63.

37. Beese, L.S., Derbyshire, V. and Steitz, T.A. (1993) Structure of DNA polymerase I Klenow fragment bound to duplex DNA. *Science* **260,** 352-355.

38. Freemont, P.S., Friedman, J.M., Beese, L.S., Sanderson, M.R. and Steitz, T.A. (1988) Cocrystal structure of an editing complex of Klenow fragment with DNA. *Proc. Natl. Acad. Sci. USA* **85,** 8924-8928.

39. Li, Y., Korolev, S. and Waksman, G. (1998) Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* **17,** 7514-7525.

40. Li, Y., Mitaxov, V. and Waksman, G. (1999) Structure-based design of Taq DNA polymerases with improved properties of dideoxynucleotide incorporation. *Proc. Natl. Acad. Sci. USA* **96,** 9491-9496.

41. Eom, S.H., Wang, J. and Steitz, T.A. (1996) Structure of Taq polymerase with DNA at the polymerase active site. *Nature* **382,** 278-281.

42. Kiefer, J.R., Mao, C., Braman, J.C. and Beese, L.S. (1998) Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature* **391,** 304-307.

43. Doublie, S., Tabor, S., Long, A.M., Richardson, C.C. and Ellenberger, T. (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* **391,** 251-258.

44. Kiefer, J.R., Mao, C., Hansen, C.J., Basehore, S.L., Hogrefe, H.H., Braman, J.C. and Beese, L.S. (1997) Crystal structure of a thermostable *Bacillus* DNA polymerase I large fragment at 2.1 Å resolution. *Structure* **5,** 95-108.

45. Tabor, S., Huber, H.E. and Richardson, C.C. (1987) *Escherichia coli* thioredoxin confers processivity on the DNA polymerase activity of the gene 5 protein of bacteriophage T7. *J. Biol. Chem.* **262,** 16212-16223.

46. Joyce, C.M. and Steitz, T.A. (1987) DNA polymerase I: from crystal structure to function via genetics. *Trends Biochem. Sci.* **12,** 288-292.

47. Catalano, C.E., Allen, D.J. and Benkovic, S.J. (1990) Interaction of *Escherichia coli* DNA polymerase I with azidoDNA and fluorescent DNA probes: identification of protein-DNA contacts. *Biochemistry* **29,** 3612-3621.

48. Polesky, A.H., Steitz, T.A., Grindley, N.D. and Joyce, C.M. (1990) Identification of residues critical for the polymerase activity of the Klenow fragment of DNA polymerase I from *Escherichia coli*. *J. Biol. Chem.* **265,** 14579-14591.

49. Turner, R.M., Jr., Grindley, N.D. and Joyce, C.M. (2003) Interaction of DNA polymerase I (Klenow fragment) with the single-stranded template beyond the site of synthesis. *Biochemistry* **42,** 2373-2385.

50. Ollis, D.L., Kline, C. and Steitz, T.A. (1985) Domain of *E. coli* DNA polymerase I showing sequence homology to *T7* DNA polymerase. *Nature* **313,** 818-819.

51. Cowart, M., Gibson, K.J., Allen, D.J. and Benkovic, S.J. (1989) DNA substrate structural requirements for the exonuclease and polymerase activities of procaryotic and phage DNA polymerases. *Biochemistry* **28,** 1975-1983.

52. Joyce, C.M. (1989) How DNA travels between the separate polymerase and 3'-5'-exonuclease sites of DNA polymerase I (Klenow fragment). *J. Biol. Chem.* **264,** 10858-10866.

53. Bailey, M.F., van der Schans, E.J. and Millar, D.P. (2004) Thermodynamic dissection of the polymerizing and editing modes of a DNA polymerase. *J. Mol. Biol.* **336,** 673-693.

54. Bailey, M.F., Van der Schans, E.J. and Millar, D.P. (2007) Dimerization of the Klenow fragment of *Escherichia coli* DNA polymerase I is linked to its mode of DNA binding. *Biochemistry* **46,** 8085-8099.

55. Carver, T.E., Jr., Hochstrasser, R.A. and Millar, D.P. (1994) Proofreading DNA: recognition of aberrant DNA termini by the Klenow fragment of DNA polymerase I. *Proc. Natl. Acad. Sci. USA* **91,** 10670-10674.

56. Bailey, M.F., Thompson, E.H. and Millar, D.P. (2001) Probing DNA polymerase fidelity mechanisms using time-resolved fluorescence anisotropy. *Methods* **25,** 62-77.

57. Datta, K., Johnson, N.P. and von Hippel, P.H. (2006) Mapping the conformation of the nucleic acid framework of the T7 RNA polymerase elongation complex in solution using low-energy CD and fluorescence spectroscopy. *J. Mol. Biol.* **360,** 800-813.

58. Johnson, N.P., Baase, W.A. and Von Hippel, P.H. (2004) Low-energy circular dichroism of 2-aminopurine dinucleotide as a probe of local conformation of DNA and RNA. *Proc. Natl. Acad. Sci. USA* **101,** 3426-3431.

59. Apetri, A.C. and Surewicz, W.K. (2003) Atypical effect of salts on the thermodynamic stability of human prion protein. *J. Biol. Chem.* **278,** 22187-22192.

60. Jha, B.K., Mitra, N., Rana, R., Surolia, A., Salunke, D.M. and Datta, K. (2004) pH and cation-induced thermodynamic stability of human hyaluronan binding protein 1 regulates its hyaluronan affinity. *J. Biol. Chem.* **279,** 23061-23072.

61. Kumar, D.P., Tiwari, A. and Bhat, R. (2004) Effect of pH on the stability and structure of yeast hexokinase A. Acidic amino acid residues in the cleft region are critical for the opening and the closing of the structure. *J. Biol. Chem.* **279,** 32093-32099.

62. Muzammil, S., Kumar, Y. and Tayyab, S. (2000) Anion-induced stabilization of human serum albumin prevents the formation of intermediate during urea denaturation. *Proteins* **40,** 29-38.

63. Nishimura, C., Uversky, V.N. and Fink, A.L. (2001) Effect of salts on the stability and folding of staphylococcal nuclease. *Biochemistry* **40,** 2113-2128.

64. Rosell, F.I., Mauk, M.R. and Mauk, A.G. (2005) pH- and metal ion-linked stability of the hemopexin-heme complex. *Biochemistry* **44,** 1872-1879.

65. Arakawa, T., Bhat, R. and Timasheff, S.N. (1990) Why preferential hydration does not always stabilize the native structure of globular proteins. *Biochemistry* **29,** 1924-1931.

66. Collins, K.D. and Washabaugh, M.W. (1985) The Hofmeister effect and the behaviour of water at interfaces. *Q. Rev. Biophys.* **18,** 323-422.

67. Goto, Y., Calciano, L.J. and Fink, A.L. (1990) Acid-induced folding of proteins. *Proc. Natl. Acad. Sci. USA* **87,** 573-577.

68. Pace, C.N. and Grimsley, G.R. (1988) Ribonuclease T1 is stabilized by cation and anion binding. *Biochemistry* **27,** 3242-3246.

69. Schellman, J.A. (1987) Selective binding and solvent denaturation. *Biopolymers* **26,** 549-559.

70. Timasheff, S.N. (1995) Solvent stabilization of protein structure. *Methods Mol. Biol.* **40,** 253-269.

71. Uversky, V.N., Segel, D.J., Doniach, S. and Fink, A.L. (1998) Association-induced folding of globular proteins. *Proc. Natl. Acad. Sci. USA* **95,** 5480-5483.

72. von Hippel, P.H. and Schleich, T. (1969) The effects of neutral salts on the structure and conformational stability of macromolecules in solution., in *Structure and Stability of Biological Macromolecules*, edited by S.N. Timasheff and G.D. Fasman, pp. 417-574, Marcel Dekker: New York.

73. Schoeffler, A.J., Joubert, A.M., Peng, F., Khan, F., Liu, C.C. and LiCata, V.J. (2004) Extreme free energy of stabilization of Taq DNA polymerase. *Proteins* **54,** 616-621.

74. Astatke, M., Grindley, N.D. and Joyce, C.M. (1995) Deoxynucleoside triphosphate and pyrophosphate binding sites in the catalytically competent ternary complex for the polymerase reaction catalyzed by DNA polymerase I (Klenow fragment). *J. Biol. Chem.* **270,** 1945-1954.

75. Sterner, R. and Liebl, W. (2001) Thermophilic adaptation of proteins. *Crit. Rev. Biochem. Mol. Biol.* **36,** 39-106.

76. Matulis, D., Kranz, J.K., Salemme, F.R. and Todd, M.J. (2005) Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor. *Biochemistry* **44,** 5258-5266.

77. Pantaoliano, M.W., Petrella, E.C., Kwasnoski, J.D., Lobanov, V.S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B.A., Lane, P. and Salemme, F.R. (2001) High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J. Biomol. Screen.* **6,** 429-440.

78. Royer, C.A. (1995) Fluorescence spectroscopy. *Methods Mol. Biol.* **40,** 65-89.

79. Warwicker, J., Ollis, D., Richards, F.M. and Steitz, T.A. (1985) Electrostatic field of the large fragment of *Escherichia coli* DNA polymerase I. *J. Mol. Biol.* **186,** 645-649.

80. Joyce, C.M. and Derbyshire, V. (1995) Purification of *Escherichia coli* DNA polymerase I and Klenow fragment. *Methods Enzymol.* **262,** 3-13.

81. Datta, K. and LiCata, V.J. (2003) Salt dependence of DNA binding by *Thermus aquaticus* and *Escherichia coli* DNA polymerases. *J. Biol. Chem.* **278,** 5694-5701.

82. Nicholls, A., Sharp, K.A. and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11,** 281-296.

83. Slater, J.P., Tamir, I., Loeb, L.A. and Mildvan, A.S. (1972) The mechanism of *Escherichia coli* deoxyribonucleic acid polymerase I. Magnetic resonance and kinetic studies of the role of metals. *J. Biol. Chem.* **247,** 6784-6794.

84. Mullen, G.P., Serpersu, E.H., Ferrin, L.J., Loeb, L.A. and Mildvan, A.S. (1990) Metal binding to DNA polymerase I, its large fragment, and two 3',5'-exonuclease mutants of the large fragment. *J. Biol. Chem.* **265,** 14327-14334.

85. Dias, A.V. and Zamble, D.B. (2005) Protease digestion analysis of *Escherichia coli* NikR: evidence for conformational stabilization with Ni(II). *J. Biol. Inorg. Chem.* **10,** 605-612.

86. Lee, Y.H., Won, H.S., Lee, M.H. and Lee, B.J. (2002) Effects of salt and nickel ion on the conformational stability of *Bacillus pasteurii* UreE. *FEBS Lett.* **522,** 135-140.

87. Melkani, G.C., Zardeneta, G. and Mendoza, J.A. (2003) The ATPase activity of GroEL is supported at high temperatures by divalent cations that stabilize its structure. *Biometals* **16,** 479-484.

88. Saboury, A.A. (2003) Application of a new method for data analysis of isothermal titration calorimetry in the interaction between human serum albumin and Ni^{2+}. *J. Chem. Thermodyn.* **35,** 1975-1981.

89. Arakawa, T. and Timasheff, S.N. (1984) Mechanism of protein salting in and salting out by divalent cation salts: balance between hydration and salt binding. *Biochemistry* **23,** 5912-5923.

90. Barnes, W.M. (1992) The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion. *Gene* **112,** 29-35.

91. Barnes, W.M. (1995) Thermostable DNA polymerase with enhanced thermostability and enhanced length and efficiency of primer extenstion. U.S. Patent 5,436,149.

92. Derbyshire, V., Freemont, P.S., Sanderson, M.R., Beese, L., Friedman, J.M., Joyce, C.M. and Steitz, T.A. (1988) Genetic and crystallographic studies of the 3',5'-exonucleolytic site of DNA polymerase I. *Science* **240,** 199-201.

93. Minkley, E.G., Jr., Leney, A.T., Bodner, J.B., Panicker, M.M. and Brown, W.E. (1984) *Escherichia coli* DNA polymerase I. Construction of a polA plasmid for amplification and an improved purification scheme. *J. Biol. Chem.* **259,** 10386-10392.

94. Datta, K. and LiCata, V.J. (2003) Thermodynamics of the binding of Thermus aquaticus DNA polymerase to primed-template DNA. *Nucleic Acids Res* **31,** 5590-5597.

95. Datta, K., Wowor, A.J., Richard, A.J. and LiCata, V.J. (2006) Temperature dependence and thermodynamics of Klenow polymerase binding to primed-template DNA. *Biophys. J.* **90,** 1739-1751.

96. Kuwamoto, S., Akiyama, S. and Fujisawa, T. (2004) Radiation damage to a protein solution, detected by synchrotron X-ray small-angle scattering: dose-related considerations and suppression by cryoprotectants. *J. Synchrotron Radiat.* **11,** 462-468.

97. Cantor, C. and Schimmel, P.R., (1980) *Biophysical Chemistry,* W.H. Freeman: San Francisco.

98. van Holde, K.E., Johnson, W.C. and Ho, P.S. (1998) *Principles of Physical Biochemistry*, Prentice Hall: Upper Saddle River, pp. 192-212.

99. Svedberg, T., Pederson, K.O. and Bauer, J.H., (1959) *The Ultracentrifuge,* Johnson Reprint Corp.: New York.

100. Philo, J. (1994) Measuring sedimentation, diffusion, and molecular weights of small molecules by direct fitting of sedimentation velocity concentration profiles, in *Modern Analytical Ultracentrifugation*, edited by T.M. Schuster and T.M. Laue, pp. 156-170, Birkhauser: Boston.

101. Philo, J.S. (1997) An improved function for fitting sedimentation velocity data for low-molecular-weight solutes. *Biophys. J.* **72,** 435-444.

102. Edelstein, S.J. and Schachman, H.K. (1967) The simultaneous determination of partial specific volumes and molecular weights with microgram quantities. *J. Biol. Chem.* **242,** 306-311.

103. Garcia De La Torre, J., Huertas, M.L. and Carrasco, B. (2000) Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* **78,** 719-730.

104. Urs, U.K., Murali, R. and Krishna Murthy, H.M. (1999) Structure of *Taq* DNA polymerase shows a new orientation for the structure-specific nuclease domain. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 1971-1977.

105. Pilz, I. (1982) Proteins, in *Small Angle X-ray Scattering*, edited by O. Glatter and O. Kratky, pp. 239-293, Academic Press Inc.: New York.

106. Feigin, L.A. and Svergun, D.I., (1987) *Structure Analysis by Small-Angle X-Ray and Neutron Scattering,* G.W. Taylor, Ed., Plenum Press: New York.

107. Svergun, D.I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76,** 2879-2886.

108. Jovin, T.M., Englund, P.T. and Bertsch, L.L. (1969) Enzymatic synthesis of deoxyribonucleic acid. XXVI. Physical and chemical studies of a homogeneous deoxyribonucleic acid polymerase. *J. Biol. Chem.* **244,** 2996-3008.

109. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. and Bairoch, A. (2005) Protein Identification and Analysis Tools on the ExPASy Server, in *The*

Proteomics Protocols Handbook, edited by J.M. Walker, pp. 571-607, Humana Press: Totowa, USA.

110. Sambrook, J. and Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd ed., Cold Spring Harbor Laboratory Press: Cold Spring Habor, Vol. 3, pp. A8.20-A28.21.

111. Aune, K.C. and Tanford, C. (1969) Thermodynamics of the denaturation of Lysozyme by guanidine hydrochloride. I. Dependence on pH at 25°. *Biochemistry* **8,** 4579-4585.

112. Blanton, T.N., Barnes, C.L. and Lelental, M. (2000) Preparation of silver behenate coatings to provide low- to mid-angle diffraction calibration. *J. Appl. Crystallogr.* **33,** 172-173.

113. Blanton, T.N., Huang, T.C., Toraya, H., Hubbard, C.R., Robie, S.B., Louer, D., Gobel, H.E., Will, G., Gilles, R. and Raftery, T. (1995) JCPDS - International Centre for Diffraction Data round robin study of silver behenate. A possible low-angle X-ray diffraction calibration standard. *Powder Diffraction* **10,** 91-95.

114. Huang, T.C., Toraya, H., Blanton, T.N. and Wu, Y. (1993) X-ray powder diffraction analysis of silver behenate, a possible low-angle diffraction standard. *J. Appl. Crystallogr.* **26,** 180-184.

115. Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H. and Svergun, D.I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36,** 1277-1282.

116. Konarev, P.V., Petoukhov, M.V., Volkov, V.V. and Svergun, D.I. (2006) ATSAS 2.1, a program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **39,** 277-286.

117. Guinier, A. (1939) La diffraction des rayons X aux tres petits angles; application a l'etude de phenomenes ultramicroscopiques. *Ann. Phys.* **12,** 166-237.

118. Guinier, A. and Fournet, G., (1955) *Small-Angle Scattering of X-rays,* John Wiley & Sons, Inc.: New York.

119. Semenyuk, A.V. and Svergun, D.I. (1991) GNOM - a program package for small-angle scattering data-processing. *J. Appl. Crystallogr.* **24,** 537-540.

120. Svergun, D.I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25,** 495-503.

121. Svergun, D.I., Barberato, C. and Koch, M.H.J. (1995) CRYSOL - a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28,** 768-773.

122. Svergun, D.I., Petoukhov, M.V. and Koch, M.H. (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80,** 2946-2953.

123. Glatter, O. (1982) Data Treatment, in *Small Angle X-ray Scattering*, edited by O. Glatter and O. Kratky, pp. 119-165, Academic Press Inc.: New York.

124. Koch, M.H., Vachette, P. and Svergun, D.I. (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q. Rev. Biophys.* **36,** 147-227.

125. Porod, G. (1982) General Theory, in *Small Angle X-ray Scattering*, edited by O. Glatter and O. Kratky, pp. 17-51, Academic Press Inc.: New York.

126. Kirz, J. (2001) Scattering of X-rays from electrons and atoms, in *X-ray Data Booklet*, 2nd edition, edited by A.C. Thompson and D. Vaughan, pp. 3-1 - 3-4, Lawrence Berkeley National Laboratory: Berkeley.

127. Krigbaum, W.R. and Kugler, F.R. (1970) Molecular conformation of egg-white lysozyme and bovine alpha-lactalbumin in solution. *Biochemistry* **9,** 1216-1223.

128. Mylonas, E. and Svergun, D.I. (2007) Accuracy of molecular mass determination of proteins in solution by small-angle x-ray scattering. *J. Appl. Crystallogr.* **40,** s245-s249.

129. Gekko, K., Kimoto, A. and Kamiyama, T. (2003) Effects of disulfide bonds on compactness of protein molecules revealed by volume, compressibility, and expansibility changes during reduction. *Biochemistry* **42,** 13746-13753.

130. Millero, F.J., Ward, G.K. and Chetirkin, P. (1976) Partial specific volume, expansibility, compressibility, and heat capacity of aqueous lysozyme solutions. *J. Biol. Chem.* **251,** 4001-4004.

131. Williams, S.P. and Langmore, J.P. (1991) Small angle x-ray scattering of chromatin. Radius and mass per unit length depend on linker length. *Biophys. J.* **59,** 606-618.

132. Zhao, J., Wang, J., Chen, D.J., Peterson, S.R. and Trewhella, J. (1999) The solution structure of the DNA double-stranded break repair protein Ku and its complex with DNA: a neutron contrast variation study. *Biochemistry* **38,** 2152-2159.

133. Osterberg, R., Sjoberg, B., Rymo, L. and Lagerkvist, U. (1975) Small-angle x-ray scattering study of the interaction between lysine transfer RNA ligase from yeast and transfer RNA. *J. Mol. Biol.* **99,** 383-393.

134. Volkov, V.V. and Svergun, D.I. (2003) Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36,** 860-864.

135. Kozin, M.B. and Svergun, D.I. (2001) Automated matching of high- and low- resolution structural models. *J. Appl. Crystallogr.* **34,** 33-41.

136. Stroud, J. (2004) Make-na server, http://structure.usc.edu/make-na/server.html.

137. Macke, T. and Case, D.A. (1998) Modeling unusual nucleic acid structures, in *Molecular Modeling of Nucleic Acids*, edited by N.B. Leontes and J. SantaLucia Jr., pp. 379-393, American Chemical Society: Washington, D.C.

138. DeLano, W.L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific: Palo Alto, http://www.pymol.org.

139. Kirschner, M.W. and Schachman, H.K. (1971) Conformational changes in proteins as measured by difference sedimentation studies. I. A technique for measuring small changes in sedimentation coefficient. *Biochemistry* **10,** 1900-1919.

140. Tanford, C. (1961) *Physical Chemistry of Macromolecules*, Wiley & Sons, Inc.: New York, pp. 364-390.

141. Karshikoff, A. and Ladenstein, R. (1998) Proteins from thermophilic and mesophilic organisms essentially do not differ in packing. *Protein Eng.* **11,** 867-872.

142. Svergun, D.I., Barberato, C., Koch, M.H., Fetler, L. and Vachette, P. (1997) Large differences are observed between the crystal and solution quaternary structures of allosteric aspartate transcarbamylase in the R state. *Proteins* **27,** 110-117.

143. Lyamichev, V., Brow, M.A. and Dahlberg, J.E. (1993) Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science* **260,** 778-783.

144. Lyamichev, V., Brow, M.A., Varvel, V.E. and Dahlberg, J.E. (1999) Comparison of the 5' nuclease activities of *Taq* DNA polymerase and its isolated nuclease domain. *Proc. Natl. Acad. Sci. USA* **96,** 6143-6148.

145. Brautigam, C.A. and Steitz, T.A. (1998) Structural principles for the inhibition of the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I by phosphorothioates. *J. Mol. Biol.* **277,** 363-377.

146. Branden, C. and Tooze, J. (1999) *Introduction to Protein Structure*, 2nd ed., Garland Publishing, Inc.: New York, pp. 383.

147. Derbyshire, V., Grindley, N.D. and Joyce, C.M. (1991) The 3'-5' exonuclease of DNA polymerase I of *Escherichia coli*: contribution of each amino acid at the active site to the reaction. *EMBO J.* **10,** 17-24.

148. Lam, W.C., Thompson, E.H., Potapova, O., Sun, X.C., Joyce, C.M. and Millar, D.P. (2002) 3'-5' exonuclease of Klenow fragment: role of amino acid residues within the single-stranded DNA binding region in exonucleolysis and duplex DNA melting. *Biochemistry* **41,** 3943-3951.

149. Bram, S. and Beeman, W.W. (1971) On the cross-section structure of deoxyribonucleic acid in solution. *J. Mol. Biol.* **55,** 311-324.

150. Luzzati, V., Masson, F. and Lerman, L.S. (1961) Interaction of DNA and proflavine: a small-angle x-ray scattering study. *J. Mol. Biol.* **3,** 634-639.

151. Qiu, X., Kwok, L.W., Park, H.Y., Lamb, J.S., Andresen, K. and Pollack, L. (2006) Measuring inter-DNA potentials in solution. *Phys. Rev. Lett.* **96,** 138101.

152. Finet, S., Vivares, D., Bonnete, F. and Tardieu, A. (2003) Controlling biomolecular crystallization by understanding the distinct effects of PEGs and salts on solubility. *Methods Enzymol.* **368,** 105-129.

153. Malfois, M., Bonnete, F., Belloni, L. and Tardieu, A. (1996) A model of attractive interactions to account for fluid-fluid phase separation of protein solutions. *J. Chem. Phys.* **105,** 3290-3300.

154. Niebuhr, M. and Koch, M.H. (2005) Effects of urea and trimethylamine-N-oxide (TMAO) on the interactions of lysozyme in solution. *Biophys. J.* **89,** 1978-1983.

155. Tardieu, A., Le Verge, A., Malfois, M., Bonnete, F., Finet, S., Ries-Kautt, M. and Belloni, L. (1999) Proteins in solution: from x-ray scattering intensities to interaction potentials. *J. Cryst. Growth* **196,** 193-203.

156. Hagerman, P.J. (1988) Flexibility of DNA. *Annu. Rev. Biophys. Biophys. Chem.* **17,** 265-286.

157. Glatter, O. (2002) The inverse scattering problem in small-angle scattering, in *Neutrons, X-rays and Light: Scattering Methods Applied to Soft Condensed Matter*, edited by P. Lindner and T. Zemb, pp. 73-102, Elsevier Science: Amsterdam.

158. Hall, C.E. (1956) Visualization of individual macromolecules with the electron microscope. *Proc. Natl. Acad. Sci. USA* **42,** 801-806.

159. Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171,** 737-738.

160. Mandelkern, M., Elias, J.G., Eden, D. and Crothers, D.M. (1981) The dimensions of DNA in solution. *J. Mol. Biol.* **152,** 153-161.

161. Petoukhov, M.V. and Svergun, D.I. (2006) Joint use of small-angle X-ray and neutron scattering to study biological macromolecules in solution. *Eur. Biophys. J.* **35,** 567-576.

APPENDIX

COPYRIGHT RELEASE PERMISSIONS

Copyright release for Chapter 2

Dear Ms Richard

We hereby grant you permission to reproduce the material detailed below at no charge in your thesis, in print and on the Louisiana State University website, subject to the following conditions:

1. If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies.

2. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier".

3. Your thesis may be submitted to your institution in either print or electronic form.

4. Reproduction of this material is confined to the purpose for which permission is hereby given.

5. This permission is granted for non-exclusive English rights only. For other languages please reapply separately for each one required.

6. This includes permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

Yours sincerely

Marion Moss
Senior Rights Assistant
Elsevier Ltd
The Boulevard
Langford Lane
Kidlington
Oxford
OX5 1GB

Tel : +44 1865 843280
Fax : +44 1865 853333
E-mail : m.moss@elsevier.com

Request From:
Allison J. Richard
Louisiana State University
107 Life Sciences Building
70803
Baton Rouge
United States

To use the following material:
ISSN/ISBN:
Title: Biochimica et Biophysica Acta
Author(s): AJ Richard et al.
Volume: 1764
Issue: 10
Year: 2006
Pages: 1546 - 1552
Article title: Thermal stability landscape for Klenow

How much of the requested material is to be used: I am requesting permission to use the full text of the article including all figures.

Are you the author: Yes

Author at institute: Yes

How/where will the requested material be used: The requested material will be used in my Ph.D. dissertation entitled "Solution Conformations and Structural Thermodynamics of Type I DNA Polymerases from Escherichia coli and Thermus aquaticus"

Copyright release for Chapter 4

Copyright Permission Policy

http://www.jbc.org/misc/Copyright_Permission.shtml

ASBMB Journals
Journal of Biological Chemistry
Molecular and Cellular Proteomics
Journal of Lipid Research
Biochemistry and Molecular Biology Education
ASBMB Today

ASBMB does not charge for and grants use without requiring your copyright permission request for:

- Original authors wanting to reproduce portions of their own work; or to republish their material in not-for-profit formats or venues.
- Students wanting to reproduce or republish their work for educational purposes.
- Students using other authors' material for their theses.
- Reproduction or republication of abstracts only.
- Photocopying up to 5 copies for personal use.
- Non-profit educational institutions making multiple photocopies of articles for classroom use; all such reproduction must utilize institutionally owned equipment for this purpose.

Use of copyrighted material requires proper citation.

VITA

Allison Joubert Richard was born in July 1979, in Eunice, Louisiana, to Michael and Sandra Joubert. Allison grew up in Opelousas, Louisiana, and attended Park Vista Elementary School. She graduated as Valedictorian from Opelousas Catholic High School in 1997. Allison began her undergraduate career in the fall of 1997 at Louisiana State University, where she majored in biochemistry and minored in classical studies. In the spring of 2001, Allison graduated *summa cum laude* with her Bachelor of Science from Louisiana State University. She entered the graduate program at Louisiana State University in the Department of Biological Sciences in the summer of 2001. As a doctoral student in the laboratory of Dr. Vince LiCata, Allison was supported by the NSF Integrated Graduate Education and Research Traineeship (NSF-IGERT) program. Allison will graduate with her doctor of philosophy degree in biochemistry on May 16, 2008 and plans to pursue a scientific career in the state of Louisiana.